

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

**Document Title: Predicting Levels of Abuse and Reassault
Among Batterer Program Participants**

Author(s): D. Alex Heckert ; Edward W. Gondolf

Document No.: 195176

Date Received: July 03, 2002

Award Number: 98-WT-VX-0014

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant final report available electronically in addition to traditional paper copies.

<p>Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.</p>

PROPERTY OF
National Criminal Justice Reference Service (NCJRS)
Box 6000
Rockville, MD 20849-6000

195176

**PREDICTING LEVELS OF ABUSE AND REASSAULT
AMONG BATTERER PROGRAM PARTICIPANTS**

National Institute of Justice, U.S. Department of Justice
(Grant No. 98-WT-VX-0014)

D. Alex Heckert
Mid-Atlantic Addiction Training Institute
and Department of Sociology
1098 Oakland Ave.
Indiana University of Pennsylvania
Indiana, PA 15705
PH: 724-357-4405
E-mail: aheckert@iup.edu

and

Edward W. Gondolf, EdD, MPH
Mid-Atlantic Addiction Training Institute
1098 Oakland Ave.
Indiana University of Pennsylvania
Indiana, PA 15705
PH: 724-357-4749; FX: 724-357-3944
E-mail: egondolf@iup.edu
Website: www.iup.edu/maati

March 8, 2002

Author's Note: The authors wish to thank the program directors at the four research sites who offered essential assistance in developing and implementing the data collection, and staff at MAATI who provided a variety of supports. We also want to thank our research assistants, especially Angie Beeman, Bob White, Jeff Rowles, and Danielle Payne. We give particular thanks to our statistical consultant, Dr. Alison Snow Jones, for her excellent statistical advice and her insightful comments on an early draft. Data coding, management, and analysis was supported by the National Institute of Justice, U.S. Department of Justice (Grant No. 98-WT-VX-0014). Data collection and parts of the qualitative analyses were supported through grants from the Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services (Grant No. R49/CCR310525). The conclusions do not necessarily represent the official view of the CDC or NIJ.

FINAL REPORT

Approved By: *[Signature]*

Date: 6/3/02

EXECUTIVE SUMMARY

Objective

Researchers, practitioners, and criminal justice personnel all recognize the need to develop better prediction of abuse and reassault among men referred to batterer programs. The prediction would help determine the extent of supervision and restraint of the batterers and assist battered women in making decisions about their safety. Currently, three main approaches have emerged in the field of domestic violence: the use of risk markers, risk assessment instruments, and batterer types. These approaches have produced relatively weak predictions for primarily dichotomized "reassault vs. no reassault" outcomes. We attempted to improve prediction by using multiple abuse outcomes and conditional factors, following the recommendations of violence researchers in other fields.

Methods

We first conducted a series of analyses using multinomial logistic regressions with multiple outcomes and conditional factors for risk markers, simulated risk instruments, and batterer personality types. The multiple outcomes included no abuse, verbal abuse or controlling behavior, threats, one reassault, and repeated reassaults during a 15-month follow-up, and conditional factors included living together, relationship troubles, antisocial behaviors, and woman filing a protection order. We then compared the results of these analyses with equations for conventional prediction with only dichotomized outcomes in order to identify any improvement in prediction. A secondary objective of our study was to explore for alternative batterers types and abuse outcome categories that might further improve prediction, or at least explain the possibility of continued weak prediction. We used a multi-site evaluation of four "well-established" batterer programs to conduct our analyses. The data set is unique because

of its large sample size (n=840), four sites, longitudinal 15-month follow-up (with interviews every 3 months), and high response rates (70% of women).

Multinomial Results

As expected, using multiple outcomes did improve prediction with risk markers. The need to distinguish "repeat reassault" from "one-time reassault" as an outcome was confirmed. Prediction was not improved, however, by including conditional variables. Nevertheless, some conditional factors did emerge as important risk markers. Future research should incorporate conditional factors, especially in explanatory or causal research, and develop conditional models to predict reassault in the short-term (e.g., over 3 months) rather than the long-term (i.e., over 12 months in this study).

We also attempted to improve prediction by simulating three popular risk assessment instruments. Only one (the Danger Assessment Scale) instrument predicted the multiple abuse outcomes as well as the risk marker analyses. Consideration of batterer personality types (based on the MCMI-III) did not significantly improve prediction. An antisocial batterer type, for instance, was not more likely to repeatedly reassault, and program format (i.e., educational vs. discussion) did not affect outcomes for different batterer types.

Exploratory Results

Our exploration of alternative conceptions of batterer types produced mixed results. We used intake risk markers to develop separate models by race. The best prediction we were able to achieve was with risk markers for whites. The prediction of repeat reassault among whites (sensitivity = 79%) modestly surpassed prediction achieved with the races combined (sensitivity = 70%). Also, batterer types based on psychopathy did not improve prediction. Batterers who repeatedly reassaulted were not significantly more likely to have evidence of psychopathy. Batterer programs were moderately successful at "teaching" some men that sanctions were

likely if they reassaulted. However, batterers perceptions of the likelihood of sanctions did not improve prediction of reassault outcomes.

Our exploration of alternative categories for abuse outcome also produced mixed results. A separate risk marker analysis for categories of non-physical abuse only marginally improved prediction of such abuse. The best predictors of non-physical abuse were previous non-physical and physical abuse; other consistent risk markers did not emerge. Qualitative analysis of violence narratives suggests that other categorizations for abuse outcome might be considered. For instance, we identified a small subset of men (4%) who engaged in unrelenting, escalating, and coercive battering was identified, but this pattern of violence was minimally associated with our "repeat reassault" outcome. Future prediction research may improve if this subcategory of battering can be consistently identified and successfully predicted.

Women's perceptions of risk were important predictors of repeat reassault throughout our multiple outcome analyses. If we can understand how women derive these perceptions, we might be able to improve other prediction efforts. We therefore attempted to identify variables associated with the women's perceptions. The strongest variables were physical and non-physical abuse, drinking behavior, and access to the partners—all of which are conventional risk markers. The women apparently rely on a constellation of abuse or a more complex process to improve their predictions.

Conclusion

Our predictive study with multiple outcomes and conditional factors partially improved prediction. Multiple outcomes improved prediction and exposed different sets of risk markers. Inclusion of conditional factors, however, did not substantially improve prediction. Interestingly, simulated risk instruments and batterer personality types did not improve prediction over the risk marker analyses. The improved prediction with multiple outcomes was modest, however.

Future research might attempt to construct more dynamic models beyond our limited use of conditional factors, and explore the process by which women assess risk.

Our findings raise implications for clinical assessment of batterers, and particularly the efforts to identify and contain the most dangerous men. The use of psychological assessments for identifying the extent of intervention or level of constraint may not be that useful in prediction. Second, risk assessment instruments appear to offer only modest prediction and should be used with caution by batterer programs and the criminal justice system. Third, it appears useful to obtain and heed women's appraisal of their situation. Fourth, "high risk" batterers may not be easily identifiable or "typed."

**PREDICTING LEVELS OF ABUSE AND REASSAULT
AMONG BATTERER PROGRAM PARTICIPANTS**

TABLE OF CONTENTS

PART I: INTRODUCTION – 1

PART II: PREVIOUS RESEARCH – 5

- INTRODUCTION – 5**
- RISK MARKERS – 5**
- RISK INSTRUMENTS – 7**
- BATTERER TYPES – 10**
- CONDITIONAL PREDICTION – 12**
- LIMITATIONS OF PREVIOUS RESEARCH – 14**

PART III: OBJECTIVES AND HYPOTHESES OF THE CURRENT RESEARCH – 16

- INTRODUCTION – 16**
- PREDICTION OF MULTIPLE OUTCOMES AND CONDITIONAL RISK MARKERS – 16**
 - Risk Markers – 17**
 - Risk Assessment Instruments – 18**
 - Batterer Types – 18**
- ADDITIONAL ANALYSES TO IMPROVE PREDICTION – 19**
 - Separate Prediction by Race – 19**
 - Batterers with Psychopathic Tendencies – 20**
 - The Effect of Men's Perceptions of Sanctions – 20**
 - Different Types of Non-Physical Abuse – 21**
 - Qualitative Analysis of Violence Narratives – 22**
 - Determinants of Women's Perceptions of Risk – 23**

PART IV: METHODS – 25

- OVERALL DESIGN – 25**
- RESEARCH SITES – 26**
- SAMPLE – 28**
- DATA COLLECTION – 29**
- MEASURES – 31**
 - Outcome Measures – 31**
 - Dichotomous Outcomes – 31**
 - Multiple Outcomes Variables – 32**
 - Missing Interviews – 33**
 - Predictors Measured at Intake – 34**
 - Social Background and Relationship Variables – 34**
 - Previous Behaviors – 34**
 - Psychological Characteristics and Batterer Types – 35**
 - Women's Characteristics – 37**
 - Program Variables – 37**
 - Predictors Measured at First 3-Month Follow-up – 38**
 - Relationship and Behavioral Variables – 39**
 - Women's Help-Seeking – 39**
 - Effects of Intervention – 39**
 - Other Interventions – 40**
 - Level of Contact – 41**

Simulated Risk Instruments – 41

Coding of Psychopathology Using the MCMI-III – 43

Qualitative Coding of Violence Narratives – 44

GENERAL ANALYTIC PROCEDURES – 45

Multinomial Logistic Regression – 45

Building the Models – 48

Nature of the Multiple Outcome Variable – 48

Intake Models – 49

Conditional Models – 50

PART V: RESULTS AND DISCUSSION FOR MULTIPLE-OUTCOME ANALYSES – 52

INTRODUCTION – 52

RISK MARKERS FOR INTAKE MODELS – 53

Dichotomous Intake Model – 53

Multiple Outcomes Intake Model – 54

Comparison of Multiple Outcomes and Dichotomous Outcomes Intake Models – 55

Risk Markers for Repeat Reassault – 56

Women's Perceptions – 57

Conclusions and Implications – 58

RISK MARKERS FOR CONDITIONAL PREDICTION MODELS – 59

PREDICTION BY RISK ASSESSMENT INSTRUMENTS – 62

BATTERER TYPES – 64

CONCLUSION – 65

PART VI: RESULTS OF FURTHER EFFORTS TO IMPROVE PREDICTION – 67

INTRODUCTION – 67

ALTERNATIVE BATTERER TYPES – 68

Risk Predictions by Race – 68

Conclusions – 69

Batterers with Psychopathic Tendencies – 70

Profile Grouping Procedure – 71

Major Findings – 72

Additional Conditional Variable: Perceptions of Sanctions – 73

Methods – 74

Procedures – 75

Results – 76

Discussion – 76

ALTERNATIVE OUTCOME CATEGORIES – 77

Risk Markers for Non-Physical Abuse – 77

Results – 78

Discussion – 79

Violent Incidents – 79

Methods – 80

Results – 81

Discussion – 82

Results for Atypical Cases – 82

Methods – 82

Results – 83

Discussion – 84

Determinants of Women's Perceptions of Risk – 85

Results – 85

Multivariate Analysis – 86

Qualitative Analysis of Women's Perceptions – 88

Discussion – 89

CONCLUSION – 89

PART VII: CONCLUSION – 91

INTRODUCTION – 91

SUMMARY OF FINDINGS – 91

LIMITATIONS OF THE CURRENT RESEARCH – 95

IMPLICATIONS FOR RESEARCH – 99

IMPLICATIONS FOR PRACTICE – 101

REFERENCES – 104

TABLES – 115

APPENDIX A – 142

PUBLICATIONS AND PRESENTATIONS SUPPORTED BY NIJ GRANT – 157

PREDICTING LEVELS OF ABUSE AND REASSAULT AMONG BATTERER PROGRAM PARTICIPANTS

PART I: INTRODUCTION

There are increased efforts among practitioners and researchers to predict reassault among men referred to batterer programs. Predicting reassault and especially "dangerous" cases has become a primary concern in the domestic violence field (Dutton et al., 1997; Goodman, Dutton, & Bennett, 2000; Hanson & Wallace-Capretta, 2000). The main reasons for this effort are to help battered women better plan for safety and to help direct limited resources toward the cases most in need of protection (Gondolf, 1997a). Accurate risk assessment is also essential in efforts to adopt and implement graduated sentencing. Under graduated sentencing, men with the lowest projected risk for reassault will receive the least amount of intervention (e.g., probation), whereas men with the highest projected risk for reassault (i.e., the most dangerous men) will receive the most severe sanctions, such as jail or increased surveillance (e.g., electronic monitoring).

Currently, three main efforts with regard to risk assessment characterize the field of domestic violence. First, researchers are attempting to identify risk markers that might help predict continuing and escalating violence. Men with markers for high risk can be targeted for more severe sanctions. In addition, the markers might identify characteristics that need to be addressed in batterer treatment to more effectively stop men's violence. For example, since heavy alcohol use is a consistent risk marker, batterer counseling may need to specifically treat alcohol use. Thus far, a number of risk markers have been identified that improve prediction of reassault beyond clinical judgment, but overall prediction is still limited and only marginally exceeds chance.

The second effort to improve prediction is the development of risk assessment instruments (Roehl & Guertin, 2000). These instruments are being developed for clinical use

and to improve the "predictive power" beyond that of risk markers. Similar instruments have been developed in related fields, such as sex offending and violence offending in general. However, the ability of these instruments to predict future domestic violence has not been sufficiently validated. The limited research to date suggests that these inventories do improve upon clinical judgment, but they still do not accurately classify men much beyond chance. It has been suggested that they will work best when used in conjunction with clinical judgment (Kropp & Hart, 2000).

The third effort to improve prediction is the identification of batterer types (Holtzworth-Munroe & Stuart, 1994; Holtzworth-Munroe, 2001). The assumption has been that there are distinct types of batterers who have different levels of risk for reassault and who may need to receive specialized types of batterer treatment (Saunders, 1996). These types have been primarily drawn from previous behavior, rather than based on their association with future behavior. There is no direct evidence, as yet, that they are predictive of continued violence.

The primary objective of our research was to improve prediction using a more complex analysis that includes both multiple outcomes and conditional factors. Multinomial logistic regression was used with a comprehensive database from a multi-site evaluation of batterer programs. We expected in the process to identify predictive risk markers, risk instruments, and batterer types. In particular, we examined the most violent and dangerous men in the sample – the repeat reassaulters – in order to uncover characteristics that distinguish them. These are the men of most programmatic and policy concern. The data set we used (Gondolf, 1997a) is particularly well suited for this effort because of the comprehensive and systematic assessment of characteristics, the apparently reliable and differentiated outcome, and the large and diverse sample of men. Previous prediction studies have suffered from shortcomings in all these areas (see Gondolf, 2001).

To accomplish our objective of improving prediction with risk markers, risk instruments, and batterer types, we performed a number of analyses. In the first stage, we tested for risk

markers in two different ways – with a static model, using factors measured at program intake, and a conditional (dynamic) model, using factors measured at program intake and at the first 3-month follow-up. We extend previous research on risk markers by using multinomial logistic regression to predict multiple outcomes – no abuse, verbal abuse/controlling behavior only, threats with no physical violence, one-time reassault, and repeat reassault – and by including conditional (dynamic) factors into the model. No research to date, to our knowledge, has developed what might be termed a “**conditional prediction model of multiple outcomes**” in the prediction of intimate partner violence. Previous research has attempted to predict a dichotomous outcome (any reassault versus no reassault), which fails to distinguish one-time reassault from repeat reassault, as well as different forms of psychological abuse. In addition, there are few studies that have estimated conditional models of reassault, which accommodate time-varying or **dynamic** independent variables, in assessing their effects on reassault (see Jones & Gondolf, 2001).

In the second stage of our research, we **simulated** three well-known domestic violence risk assessment instruments to assess how well they predict repeat reassault and whether they predict better than the risk markers identified in the first stage. We do this by regressing the multiple outcomes on each of the risk instruments separately and comparing these models to the previous models.

In the third stage, we assessed whether prediction can be improved by consideration of “batterer types,” as measured using a personality test administered in the evaluation. We did this by determining whether batterer type is predictive of reassault after controlling for other significant risk markers. We also examined the interaction of “batterer type” with program format to assess whether different types of batterers would benefit from different types of formats (i.e., instructional versus discussion).

A secondary objective of our study is to explore additional ways to improve prediction.

In this stage of our research, we conducted a variety of quantitative and qualitative analyses on

alternative subtypes of batterers and alternative categories of abuse outcomes. First, we developed separate prediction equations for the three main racial groups in our data set – African Americans, Whites, and Latinos – to determine if significant risk markers vary across racial groups. Second, we examined the relationship of psychopathy and psychopathic tendencies, based on the responses to the MCMI-III profiles (Gondolf & White, 2001), to abuse outcomes. The refined measures of psychopathology may improve prediction of abuse outcomes considering the criminological research linking psychopathy and violence. Third, we explored an additional conditional factor—the impact of deterrence on prediction. We entered the men’s perceptions of the likelihood of sanctions as a conditional factor for reassault (Heckert & Gondolf, 2000a). Fourth, we expanded the categories for non-physical abuse and develop separate predictive equations for these categories (Gondolf, Heckert, & Kimmel, in press). These equations help refine the risk markers and prediction for emotional abuse separately from reassault. Fifth, we explored additional outcome categories through qualitative analyses of violence narratives (Gondolf & Beeman, under review) and atypical cases. These analyses indicate whether different conceptions of outcome might ultimately improve prediction. Sixth, because women’s perceptions of risk proved to be predictive of repeat reassault in our models, we developed equations to identify determinants of these perceptions (Heckert & Gondolf, 2001). Perhaps actuarial prediction of batterer’s risk of reassault can be improved by understanding how their partners develop perceptions of risk.

PART II: PREVIOUS RESEARCH

INTRODUCTION

Prediction efforts include lists of predictive factors or risk markers (see Saunders, 1995; Dutton & Kropp, 2000; Hanson & Wallace-Capretta, 2000), batterer profiles and types (see Holtzworth-Munroe & Stuart, 1994), and risk assessment scales or indexes (see Dutton & Kropp, 2000; Roehl & Guertin, 2000). We will review briefly the research in each of these areas to illustrate the advances and limitations in the effort to predict violence, particularly within clinical samples. Two points stand out. One, the predictive power has been relatively weak. Two, prediction is likely to be improved with more complex analyses – specifically those that account for multiple outcomes, rather than simplistic dichotomous “success-failure,” and for additional conditional factors (i.e., “time-varying” circumstances or conditions). Such analyses require, however, large, comprehensive, and longitudinal data sets. We also consider the logical implication of what might be considered conditional prediction of multiple outcomes – that is “on-going risk management” as an alternative to the more conventional static risk assessments, based primarily on information obtained at program intake. On-going risk management involves assessing and reassessing risk periodically in response to incremental intervention and changing circumstances.

RISK MARKERS

Numerous studies have attempted to identify risk markers for reassault (see Saunders, 1995; Dutton & Kropp, 2000). By “risk markers” we mean characteristics that increase the likelihood of reassault. If a person has one of these characteristics, the likelihood of reassault increases. On the other hand, a risk marker does not mean that a particular person is “going” to reassault or should necessarily be classified as a “high risk” case. Risk markers are not necessarily “causes” for reassault (although they might be). They are simply characteristics that

are associated with reassault. For example, unemployment may be a risk marker for violence, but it is not unemployment that causes the violence. The financial consequences of unemployment may create stress within the individual and between a couple, and unemployment could mean that the man is at home with his partner more or drinking more, all of which may increase the risk of reassault.

In risk prediction studies, the risk markers are typically drawn from background information obtained at program intake. They are generally characteristics that are readily observable and accessible to practitioners (or clinicians), rather than the results of extensive testing and scales. The objective is to derive risk markers that practitioners can easily recognize and apply, and thus use to improve their clinical assessments and judgments.

A number of risk markers have been identified. Goldkamp (1996) found that prior criminal behavior was associated with reassault. Indications were that drug involvement (not alcohol), prior convictions, and prior assault and/or battery arrests interacted to produce greater risk of rearrest within seven months for a domestic violence offense. A longitudinal analysis with a multi-site evaluation data set found men who were reportedly "drunk" during the follow-up period were substantially more likely to reassault their partners (Jones & Gondolf, 2001). Batterer program status has also been associated with reassault. Program dropouts are significantly more likely to reassault (Dutton, 1986; Edleson & Syers, 1990; Gondolf, 1997b; Hamberger & Hastings, 1988), even when controlling for demographic and behavioral differences. "Voluntary" program participants are also more likely to reassault than court-mandated participants (Gondolf, 1997b). According to a large, multi-site evaluation, length of program and program format (instructional versus discussion) do not appear to be predictive of reassault rates, even when controlling for other risk markers (Gondolf, 1999a). One small experimental study, however, suggests that program format (i.e., cognitive behavioral versus psychodynamic) may interact with personality type (i.e., antisocial versus depressive) (Saunders, 1996).

Across studies, some fairly consistent risk markers have been identified, such as prior assault, excessive alcohol or drug abuse, previous criminality, severe personality disorders and/or psychological problems, neglect or abuse as a child, and program dropout (see DeMaris & Jackson, 1987; Dutton, et al., 1997; Gondolf, 1997b; Hamberger & Hastings, 1990; Hilton, Harris, & Rice, 2001; Saunders, 1995; Tolman & Bennett, 1990). Nevertheless, the predictive power of these factors, even when combined together, is weak, with a limited ability to correctly classify reassault on a better than chance basis (Limandri & Sheridan, 1995; Saunders, 1995). The risk markers incorrectly identify a lot of men. Many men who do not reassault are predicted to be "reassaulters" or at "high risk" for reassault (false positives), and many men who do reassault are predicted to be "non-reassaulters" or at "low risk" for reassault (false negatives).

RISK INSTRUMENTS

A concurrent development in the domestic violence field is the use of risk assessment inventories, such as the Spousal Assault Risk Assessment (SARA), the Domestic Violence Inventory (DVI), and the Kingston Screening Instrument for Domestic Violence Offenders (K-SID). These instruments have expanded upon the long-standing effort on the part of practitioners to develop lethality checklists, such as the Danger Assessment Scale (DAS) (Hart, 1994). Risk instruments have also been widely used in assessing sex offenders and violent offenders in general (e.g., Sex Offender Risk Appraisal Guide, SORAG, and Violence Risk Appraisal Guide, VARG; Quinsey, Harris, Rice, & Cormier, 1998). All the instruments attempt to improve the "predictive power" beyond that of risk markers by assembling a constellation of risk markers. The instruments measure several or many risk markers and calculate a score that reflects the degree of risk. The instruments may be self-administered (i.e., the man answers the items himself) or clinically-administered (i.e., a clinician rates the man based on interviews and records). The latter approach can also incorporate information gathered from the victim. The

clinical approach obviously takes more time and expertise, and relies on cooperation from the women and the availability of accurate records.

One challenge with these instruments, and prediction in general, is obtaining the needed and accurate information. The men tend to inaccurately respond because of concern about the criminal justice consequences (Heckert & Gondolf, 2000b). Will they be put in jail if they tell the whole truth about their violence? On the other hand, women are often reluctant to respond because of fear of retaliation, concerns about their relationship, or because they don't know exactly what will happen (Heckert & Gondolf, 2000b, 2000c). A jurisdiction recently experimenting with a clinically-administered instrument was able to contact only a third of the female partners (Houghton, 1999).

The risk assessment instruments for intimate partner violence have not been sufficiently validated (based on predictive validity) because of a limited number of follow-up studies and shortcomings of the existing validation studies (Healey et al., 1998; Roehl & Guertin, 2000). To date, risk assessment inventories offer at best marginal prediction (Dutton & Kropp, 2000; Roehl & Guertin, 2000). The inventories do, however, offer an improvement over clinical judgment and appear to have more predictive power than a clinical application of risk markers (i.e., a clinical assessment guided by risk markers) (Dutton & Kropp, 2000; Roehl & Guertin, 2000). It has been argued that they will work best when used to enhance clinical judgment, which has been supported by one study (Kropp & Hart, 2000). The clinical utility of such instruments as prediction tools, however, remains debatable. They do not appear to correctly classify men at a clinically acceptable level, as the instruments still make a substantial amount of misclassifications, just as in the research on risk markers (e.g., Kropp & Hart, 2000).

Development of the risk assessment instruments and more sophisticated prediction research continues. Hanson and Wallace-Capretta (2000), for example, found that a structured risk measure designed for general offenders, the Level of Service Inventory – Revised (LSI-R), and age were the only multivariate predictors of violence recidivism (as measured by arrests) in

a Canadian sample. The LSI-R is based on social learning theory and includes subscales pertaining to criminal history, problems with education and employment, finances, relationships (including attitudes about women, partner assault, and service providers), unstable housing patterns, leisure activities, and substance abuse. Because most of this violence recidivism was domestic violence, Hanson and Wallace-Capretta contend that the LSI-R should be used as a risk instrument for domestic violence unless future research demonstrates the superiority of instruments developed specifically for predicting domestic violence.

A recent study (Hilton, Harris, & Rice, 2001) found that the Violence Risk Appraisal Guide (VRAG) was a slightly better, although modest, predictor of violence recidivism (not just partner reassault) among eighty-eight serious wife assaulters ($r = .42$) than the Hare Psychopathy Checklist-Revised ($r = .35$). The VRAG is a twelve-item risk assessment instrument used to predict violent recidivism. It includes static demographic variables, childhood history, criminal history, and psychiatric assessment variables. Like Hanson and Wallace-Capretta (2000), Hilton and her colleagues (2001) question whether a specialized risk assessment instrument is necessary for predicting future violence among wife assaulters. Nevertheless, their results are limited by the small sample size ($n = 88$) and the extreme and homogeneous nature of their sample of wife assaulters. All the subjects had been admitted to a maximum-security psychiatric facility and over half (59%) had an index offense of murder or manslaughter.

Several cautions have emerged in the reviews of instruments used to predict men's risk of reassault. The reviews recommend that cut scores should not be used as the main means for determining a batterer's risk (Dutton & Kropp, 2000). In addition, scoring should be used in conjunction with other information and with advice from battered women's advocates. The instruments' results should not be used to pressure women to accept a certain sentence for her partner, or to justify substandard supervision or treatment of arrested men. The main benefit of the instruments may be in the awareness and communication they encourage (Roehl & Guertin,

2000). Simply getting police, court officials, women's advocates, and batterer counseling to routinely discuss possible risk may do much to reduce risks.

BATTERER TYPES

The related emphasis in practice and research is to identify specific types of batterers (e.g., through batterer typologies), especially high-risk offenders, in an effort to improve prediction and develop specialized interventions to accommodate these different types of offenders (see Healey, et al., 1998; Saunders, 1996). Rather than a host of risk markers, men are sorted into categories or "types" based on characteristics that are highly associated with one another. Each type implies a unique set of treatment needs and a likely outcome of its own. The types may give a more meaningful and useful portrayal of batterers than lumping all the men together to predict outcome, as is done in deriving risk markers.

Several studies of batterer types have used personality characteristics (e.g., Hamberger et al., 1996) or behavioral patterns (e.g., Gondolf, 1988). Some have drawn on clinical assessments, but most have statistically clustered the characteristics (e.g. Hamberger et al., 1996). These types have used reports of previous violence (i.e., prior to program intake) to verify that some types tend to be more violent than others. A major review summarized this research with three prevailing types of batterers: family-violent only, dysphoric/borderline, and violent/antisocial (Holtzworth-Munroe & Stuart, 1994).

A study using a prominent personality inventory, the MCMI-I (Millon Clinical Multiaxial Inventory, Version I) across two sites, similarly, identified three major types of men according to a clustering of the subscale scores (Hamberger et al., 1996). It found non-pathological, passive-aggressive/dependent, and antisocial types. A laboratory study of recruited batterer couples (n=60) identified two major types based on physiological cues and observed behaviors: angry "pit bulls" and cold-hearted "cobras" (Jacobson & Gottman, 1998). In another study, a combination of a personality test and clinical assessment suggested one prevailing type among

batterer program participants termed the "abusive personality." Similar to the "pit bulls," these men were easily frustrated and angered largely as a result of their borderline personality problems (Dutton, 1998; Dutton & Starzomski, 1993; Tweed & Dutton, 1998). All of these typologies appear to consider an impulsive, angry batterer (i.e., dysphoric/borderline) and a more deliberate, generally-violent man (i.e., violent/antisocial), much like the classic conception of "expressive" and "instrumental" violence (Berkowitz, 1993).

A recent study investigated the potential of personality-based "batterer types" using MCMI-III profiles of batterers in a multi-site evaluation (White & Gondolf, 2000). Each profile represents the combination of scores a man received on the MCMI subscales. Following the procedures outlined in the MCMI interpretive manuals, White and Gondolf characterized and then grouped the profiles in terms of dysfunction, predominant personality tendency, and prevalent commonalities. Their profile interpretations produced three levels of personality pathology referred to as low personality dysfunction (56%), moderate personality dysfunction (29%), and severe personality dysfunction (15%). The profile groupings loosely reflected the "types" derived from the previous research. The "low" dysfunction narcissistic/conforming style is comparable to the "family-violent only" type asserted in the landmark research review (Holtzworth-Munroe & Stuart, 1994). The avoidant/depressive style and borderline disorder reflect the "dysphoric/borderline" type, and the antisocial, narcissistic, and paranoid disorders are similar to the "generally-violent/antisocial" type. The profile groupings also approximate the types derived from a cluster analysis of the MCMI-I subscales (Hamberger et al., 1996).

The identification of types is consistent with patient-treatment matching trends in clinical psychology and alcohol treatment (Beutler & Clarkin, 1990; Gordis, 1997). As mentioned earlier, a clinical trial of batterer counseling approaches (n=218) found that men with an antisocial diagnosis were less likely to reassault their partners if they completed a gender-based, cognitive-behavioral group (Saunders, 1996). Men with a dependent diagnosis had better outcomes if they participated in a process-psychodynamic group. The "diagnoses" were

based on scores from the MCMI subscales (BR > 74). The implication is that the "one size fits all" approach of many batterer programs is inappropriate and may be counterproductive (Geffner, 1995).

In summary, current typologies are based either on psychological characteristics (Dutton & Golant, 1995; Hamberger et al., 1996; White & Gondolf, 2000) or criminal justice or behavioral factors such as demographic information, criminal histories, and substance abuse data (Goldkamp, 1996; Gondolf, 1988). With the exception of Goldkamp's criminal justice-based typology, which was predictive of rearrest for domestic violence, these typologies have not yet been confirmed as predictive of different levels of abuse among men arrested and sent to batterer counseling. It is also unclear whether "patient-treatment matching" will improve treatment outcomes and thus prediction of high-risk offenders. An extensive study of patient-treatment matching for individuals addicted to alcohol showed minimal improvements in outcomes (Gordis, 1997).

CONDITIONAL PREDICTION

An alternate trend is the effort to identify conditional or dynamic predictors of reassault (Hanson & Wallace-Copreta, 2000; Milner, 2001). Most previous research has used static variables measured prior to treatment, such as previous criminality and abuse, education, and personality type. Risk assessment based on static predictors is easier to implement into criminal justice policy and may be suitable for long-term prediction of reassault. However, incorporation of time-varying or dynamic factors may not only improve long-term prediction, but also may facilitate the prediction of the timing of repeat reassaults, as well as changes in the level of risk. With regard to policy, conditional risk models can be translated into "ongoing case management," whereby the criminal justice system is responsive to changes in batterers' reactions to subsequent intervention and their modified likelihoods of reassault. Such proactive policies may therefore reduce the level of risk to potential victims.

In conditional prediction, violence is viewed as a process – an interaction between individual characteristics and varying situations or circumstances (i.e., conditions). Most risk assessment instruments do not account for this process and the situations it involves. An array of studies sponsored by the MacArthur Foundation examined and promoted the contribution of conditional treatment with violent psychiatric patients (Monahan & Steadman, 1994). Predictors for violence among psychiatric patients confirmed that short-term predictions are fairly reliable (Mulvey & Lidz, 1993). These predictions are generally an extrapolation of current violence or hostility, rather than a prediction of future violence. For longer-term predictions, conditions after treatment matter most – conditions such as access to family members, compliance with prescribed medications, and participation in outpatient counseling.

A recent study of potential assassins by the U.S. Secret Service (Fein, Vossekuil, & Holden, 1995) furthers this view. It acknowledges obvious predisposing factors such as a history of “problems,” escalating threats, and an obsession with the potential victim, but it is critical of static risk instruments. The Secret Service study recommends that assessment interviews explore the plans, circumstances, and thresholds of violence; and do re-assessment based on these types of situational factors. A threatening individual, then, needs to be monitored for changes in these situations.

Both reports recommend on-going risk management and “target hardening” to account for the situations. On-going risk management entails periodic monitoring of an individual’s situation. Is he taking his medication, living alone, stopped stalking, and not drinking heavily? If there is evidence of any changes that would lead to violence, an appropriate intervention is made. If the man is drinking, he might be sent back to alcohol treatment. The man is then monitored again to see if he is complying with the new intervention and if any other changes have occurred.

“Target hardening” is the other side of monitoring the individual. It refers to the increased awareness, support, and protection given to potential victims. A potential victim might

be made aware of certain "warning signs" or cues, and have security surveillance intensified. The victim is often in the best position to recognize the escalation of threats and take preventive action.

LIMITATIONS OF PREVIOUS RESEARCH

The prediction research has, in general, been limited by several factors. Most findings are based on small samples of less than 100 subjects, studies with low response rates of under 50%, a single follow-up interview, program completers only, single-site designs, and follow-up reports from only the initial victim (Gondolf, 1997a). Saunders (1996) documents the impact of partner contact and new partner reports on reassault rates. The previous research has also relied on simplistic dichotomous outcomes of "success" and "failure," not considered dynamic or "conditional" variables, and databases have been too small or limited to address these problems (Monahan, 1996; Mulvey & Lidz, 1993). Prediction is further hampered by dubious reliability because of long recall periods, and uncertain validity because there is no contact with new partners or search of police records. The previous research designs do not, moreover, sufficiently distinguish men who repeatedly reassault and are the most dangerous. These men appear to warrant the most extensive intervention.

Prediction in the domestic violence field may, moreover, be compromised by conceptual and data limitations made apparent by advances in research on violent psychiatric patients. Researchers in the psychiatric field have demonstrated the utility of differentiating outcomes rather than relying on a dichotomized variable of "reassault" versus "no reassault" (Monahan, 1984; Mulvey & Lidz, 1993; Steadman et al., 1994). The effects of predictors become more apparent with a more sensitive outcome variable. A 15-month multi-site evaluation of batterer programs (Gondolf, 1997b) shows rates of verbal abuse and threats (70%) to be much higher than reassault rates (35%) and the severity of reassault patterns to vary substantially between repetitive reassault (23%) and episodic reassault (12%). This range of abuses in itself suggests

the importance of broadening outcomes in prediction efforts in the domestic violence field. Although verbal abuse is illegal only in extreme forms of harassment or threats, it falls under most definitions of domestic violence and may be a part of a pattern of abuse, which many batterer interventions are attempting to change (Healey et al., 1998).

As suggested above, psychiatric researchers have additionally improved prediction by introducing situational and intervening (dynamic) variables to formulate models of "conditional prediction" (Lidz et al., 1993; Mulvey & Lidz, 1985). Conditional prediction accounts for a variety of intervening or time-varying factors, such as whether a psychiatric patient takes his or her medication, the access a patient may have to potential victims, and surveillance in terms of support groups, outpatient care, and additional hospitalization. In the case of batterers, these "conditions" (or intervening variables) may translate into program participation, access to one's partner, and additional interventions and services to the batterer and his partner – after program intake.

Large, complex, longitudinal data sets are, however, required to develop and test what amounts to a conditional prediction of multiple outcomes. As stated earlier, previous data sets of batterers have been based on small samples of approximately 100-200 subjects with response rates of approximately 50% for typically a single follow-up interview (Gondolf, 1997a; Tolman & Bennet, 1990). These data sets are not large enough to offer the statistical power to assess multiple outcomes and conditional variables. Consequently, developing a more sophisticated prediction model of batterer reassault has been limited by a lack of a suitable database, as well as simplistic conceptions of outcomes.

PART III: OBJECTIVES AND HYPOTHESES OF THE CURRENT RESEARCH

INTRODUCTION

In this study, we examine the prediction of abuse and assault among batterer program participants by addressing the methodological shortcomings outlined above. We attempt to improve prediction of especially reassault using multiple outcomes and conditional variables. To do so, we use an extensive, multi-site longitudinal database of batterers and their female partners to test the contribution of conditional prediction with multiple outcomes using risk markers, risk assessment instruments, and batterer types. The database also allows us to explore alternative batterer types and outcome categories. It includes intake interviews with 840 batterers and their partners and follow-up interviews every 3 months over 15 months, with a 70% response rate over the follow-ups.

In this part of the report, we describe our primary objectives, which are to assess whether prediction of reassault can be improved by examining multiple outcomes, considering conditional predictors, using risk instruments, and considering batterer types. We then discuss secondary objectives that include additional efforts to improve prediction by considering additional batterer types and abuse categories.

PREDICTION OF MULTIPLE OUTCOMES AND CONDITIONAL RISK MARKERS

The primary objective of our proposed research is to test a "conditional prediction model of multiple outcomes" of batterer intervention using multinomial (i.e., polytomous) logistic regression. The main contribution of the model is that multiple outcomes are considered rather than simply a dichotomous "success" or "failure". We use batterer and partner self-reports of physical assault, verbal abuse, controlling, and threatening behaviors to identify 5 distinct outcomes for batterer program participants: **non-abusive behaviors during follow-up** (19% of the sample), **controlling behavior/verbal abuse only** (26%), **threatening assault with no**

physical reassault (20%), one-time reassault (12%), and repeat reassault (23%). This latter type of abuse is highly associated with severity and injury, and presumably dangerousness.

Risk Markers

Our first goal is to determine if this more sophisticated approach to prediction improves prediction with risk markers. We used multinomial logistic regression to determine which static risk markers measured at intake are most predictive of the multiple outcomes. We also determined which static risk markers are most predictive of a dichotomous outcome (no reassault versus reassault) for comparison. The comparison of these two sets of "markers" allowed us to assess whether multiple outcomes improves prediction of batterers' reassaultive behavior. We hypothesized that the multiple outcomes for abuse would improve prediction and identify additional risk markers.

We separately considered the influence of conditional or situational variables that occur after program intake. These variables were assessed at three months following program intake and entered into regression equations for multiple abuse outcomes after risk markers, batterer types, and program variables assessed at intake. The "conditional" variables included batterer or victim employment, partner contact and new partners, batterer's perceptions of sanctions for program dropout and reassault, alcohol and drug use, batterer alcohol or psychological treatment, the women's use of victim services and other help sources, and additional criminal justice intervention. Selection of these variables as predictors was guided by previous theory and research (Cano & Vivian, 2001; Dutton et al., 1997; Gendreau, Little, & Goggin, 1996; Hanson & Wallace-Capretta, 2000; Heckert & Gondolf, 2000; Hilton, Harris, & Rice, 2001; Jones & Gondolf, 2001; Schumacher, Feldbau-Kohn, Slep, & Heyman, 2001). The model is hierarchical in that the effects of intake variables on program outcomes are assessed first. The effects of the intervening conditional variables are assessed in a second step, after controlling

for the effects of the intake variables. We hypothesized that prediction of multiple outcomes would be improved by including these conditional variables in the multivariate models.

Risk Assessment Instruments

A second goal was to examine the prediction of risk assessment inventories using our prediction model of multiple outcomes. We **simulated** several popular risk assessment instruments – the Kingston Screening Instrument for Domestic Violence Offenders (K-SID), Campbell's Danger Assessment Scale (DAS), and the Spousal Assault Risk Assessment instrument (SARA) – using variables measured at intake. We computed logistic regression equations in which we regressed a dichotomous outcome on the three simulated risk instruments separately. For comparison, we then computed the same equations for multiple outcomes using multinomial logistic regression. We hypothesized that multiple outcomes would modestly improve the predictive power of the risk assessment instruments. We also hypothesized that the instruments would not provide as accurate prediction as the model using intake risk markers and conditional variables.

Batterer Types

Our third goal was to examine whether batterer types, as measured by the MCMI-III, are predictive of multiple outcomes, when controlling for other significant risk markers. We also ascertained whether prediction of batterer program outcomes could be improved by considering batterer type in conjunction with program format (i.e., instructional versus discussion). It has been argued that different types of batterers require different types and levels of treatment, which can be assessed in a preliminary fashion by constructing an interaction term between batterer type and program format to see whether it makes a significant contribution to the prediction model. We hypothesized that batterer types – specifically an “antisocial” type – would be a significant risk marker for repeat reassault. We also hypothesized that the “antisocial”

batterer type would interact with program format to significantly improve prediction of reassault outcomes.

ADDITIONAL ANALYSES TO IMPROVE PREDICTION

A second major objective of our research was to explore for alternative ways to improve prediction. There may be alternative subtypes of batterers or categories of abuse to the ones examined by our conditional model of multiple outcomes. To explore for alternatives, we drew on information in the data set that addresses alternative explanations. For instance, clinicians are increasingly arguing that racial background may lead to different prediction models for program outcomes (see Gondolf & Williams, 2001). Also, the process of violence may suggest a more complex set of abuse outcomes than the ones we initially employed (see Monahan, 1996; Mulvey & Lidz, 1993; Steadman, 1982).

Separate Prediction By Race

We first developed separate prediction equations for each of the three main racial groups represented in the data set: African American, White, and Latino. The geographical distribution of the sample conveniently provided a fairly even distribution of the three groups necessary for such an analysis. There has been a tendency to over-generalize from prediction research on Whites in the field of domestic violence. Race has typically been ignored, used as a control variable, or examined as a separate risk factor. It is possible that there are different risk markers for reassault among the three racial groups (see Rankin, Saunders, & Williams, 2000). To identify these different risk markers, we developed multinomial logistic regression models for each racial group. However, owing to a smaller sample size for Latinos, a binary logistic regression model (i.e., dichotomous outcome) was developed for Latinos. Because of sample size constraints, we were also unable to develop conditional models, using predictors measured after intake, for any of the three racial groups. We hypothesized that different risk

markers for reassault would be identified across the three racial groups and prediction would be improved over prediction with the combined samples.

Batterers with Psychopathic Tendencies

We also examined other subtypes of batterers. First, we identified “primary” and “secondary” psychopathy using the MCMI-III profiles (see Gondolf & White, 2001). We also identified batterers with any “psychopathic tendencies.” We used all three categorizations of psychopathy as subtypes for predicting multiple outcomes. We examined whether primary psychopathy, secondary psychopathy, or psychopathic tendencies were predictive of repeat reassault, both at the bivariate level and with multivariate controls. In addition, we assessed the predictive ability of these measures of psychopathology in each of the three main racial groups in our study. We hypothesized that batterer subtypes with primary psychopathy, secondary psychopathy, and psychopathic tendencies would be more likely to reassault, and especially to repeatedly reassault – overall, and within racial groups.

The Effect of Men's Perceptions of Sanctions

We addressed an additional conditional factor, as well. We explored how the perception of sanctions affected program dropout and reassault. Deterrence theory would argue that men who perceive sanctions as likely are often deterred from further violence. We attempt, therefore, to test the influence of perceived sanctions on prediction.

A persistent theme in the intervention with men who batter women has been the deterrent effect of severe and certain sanctions. The likelihood of jailing, for instance, supposedly helps motivate “batterers” to comply with batterer programs and desist from further assault. Several demonstration projects have attempted to coordinate the criminal justice response with batterer programs to ensure a swift and sure punishment for program dropout or

reassault (National Institute of Justice, 1999). These efforts reflect the goals of the movement toward a "coordinated community response," currently being promoted in the field (Hart, 1995).

While a great deal of research has been conducted regarding deterrence, relatively few studies have examined the specific deterrence effects of legal sanctions in the field of domestic violence. We therefore examined the extent, nature, and effect of the batterer subtype – men who perceive sanctions (jail) as likely – on program dropout and partner reassault. First, we examined whether program context, such as mandatory court review of compliance, cause batterers to perceive sanctions as likely. Second, we identified other predictors for this subtype to explore for an "experiential effect" on perceptions (i.e., experience with the system influences perceptions). Third, we analyzed whether the subtype of batterer who perceives sanctions as certain and severe will be less likely to dropout of the program and to reassault over a 15-month follow-up. This analysis offers a test of "specific deterrence" (i.e., the direct preventive effects of legal sanctions imposed on individuals who actually engage in illegal behavior). Based on the assumptions of coordinated criminal justice intervention and the existing domestic violence research, we hypothesized that we would find a program, experiential, and specific deterrence effect on reassault and program dropout. In other words, men who stayed in the program (i.e., did not dropout early) and who had previous contact with social control agencies would be more likely to perceive certain and severe sanctions, and these men would be less likely to dropout later in the program or to reassault.

Different Types of Non-Physical Abuse

We explored the possibility of improving prediction of at least some forms of abuse by examining more fully alternative categories of non-physical abuse. These categories were somewhat neglected in our initial set of multiple outcomes, which distinguished only threats versus the other forms of non-physical abuse together and which focused on distinguishing repeat reassault. We identified risk markers separately for four categories of non-physical

abuse reported by the batterers' female partners during follow-up: controlling behavior, verbal abuse, threats, and severe non-physical abuse. Non-physical abuse of women is receiving increasing attention because of its effect in diminishing another person's self-esteem and marital well-being (O'Leary, 1999). Moreover, batterer counseling programs generally link non-physical and physical abuse as part of a broader pattern of men's effort to exert power and control over their female partners (e.g., Pence & Paymer, 1993). Predictive research on non-physical abuse has primarily focused on identifying predictors among general population samples (Follingstad et al., 1990; Margolin, John, & Foo, 1998; Stets, 1991; Sugarman, Aldarondo, & Boney-McCoy, 1996), while research with clinical samples has not addressed the risk markers for non-physical abuse. We hypothesized that the additional categories for non-physical abuse outcomes will reveal additional risk markers and improve prediction over the more broad categorization of non-physical abuse.

Qualitative Analysis of Violence Narratives

We also explored for the possibility of alternative categories or categorizations of violence. These alternatives might help explain the weak predictions using conventional behavioral-based categories of reassault. Some researchers have argued that violence must be interpreted as a process, and that the process exposes patterns and dynamics beyond the incident's behavioral act (e.g., push, hit, etc.) (Monahan, 1996; Mulvey & Lidz, 1993; Steadman, 1982). The women's narratives of reassault were analyzed to identify possible alternative abuse outcomes using a sequential, situational model of violence suggested in the process conception of violence. We hypothesized that alternative categories for violence may be evident. Specifically, men who repeatedly reassault may not necessarily account for an excessive, unrelenting, and escalating pattern of violence.

We also conducted a qualitative analysis to identify atypical categories of batterer "types" and abuse outcomes that might have influenced our ability to predict reassault. There is

more diversity among the men who batter and their female partners than the categorical measures of characteristics and statistical equations can represent. Many men and women face unique circumstances that exceed the conventional measures of demographics, personality, and past behavior. Descriptive narratives from approximately 120 "atypical" cases were summarized by two graduate assistants and then grouped according to the distinguishing features of the case. We hypothesized that the "atypical cases" would identify diversities and categorizations that could reduce prediction using conventional batterer types and abuse outcomes.

Determinants of Women's Perceptions of Risk

A final area warranting further exploration emerged from our risk assessment analyses. As we will show in the results section, women's perceptions of how likely it is the man will use violence in the future and how safe she feels improved the prediction of threats and reassault (including repeat reassault), even when controlling for other risk factors. This finding reinforces recent research that showed women's perceptions of safety help predict the likelihood of reassault over four months, while controlling for other predictors (Weisz, Tolman, & Saunders, 2000). Identification of determinants of women's perceptions could, therefore, help researchers and practitioners improve their prediction of reassault. Langford (1996), in his theory of "predicting unpredictability," argued that women simultaneously monitor and respond to cues in their partners and pursue strategies to avoid abuse. Because women often use subtle cues and situational cues (Langford, 1996; Stuart & Campbell, 1989), because of the dynamic nature of abusive relationships, and because women pursue various strategies in response to danger cues (Langford, 1996), it may be difficult to develop strong explanatory models of women's perceptions or risk appraisals. Nevertheless, it is likely that women do attend to many of the risk factors identified by prevailing risk prediction research when assessing risk. Thus, we hypothesized that the best predictors of women's perceptions would be relationship status.

drinking behavior, and past physical and psychological abuse. We also hypothesized that, overall, the explanatory models of women's perceptions will be relatively modest.

PART IV: METHODS

OVERALL DESIGN

To answer the research questions, we used a multi-site database of batterers and their female partners. This database includes 840 men who were admitted to batterer programs in four cities – Pittsburgh, Dallas, Houston, and Denver. Interviews were conducted at program intake with batterers, initial victims, and new female partners every 3 months. We used data from the intake interviews, as well as from the first five follow-ups, giving us 15 months of follow-up. To test our primary objectives, we used multinomial logistic regression to develop equations to predict multiple outcomes, which we compared to binary logistic regression equations used to predict a dichotomous outcome. We built intake equations, which used risk markers measured at intake, as well as conditional equations, which used risk markers measured at intake and at the first 3-month follow-up. We also used multinomial logistic regression to assess the level of prediction obtained by risk instruments, women's perceptions of risk, and batterer types. To accomplish our secondary objectives, we conducted further explorations of alternate batterer types and abuse categories using the rich qualitative and quantitative information available in the data set.

In this part of the report we first describe the research sites, and then describe the sample strategy and sample characteristics. Next, we describe the data collection methods, which is followed by descriptions of the variables used in our analyses. We then discuss the general analytic procedures, and specifically, we explain why multinomial logistic regression is the appropriate technique for analyzing a dependent variable with multiple outcomes. Finally, we describe how we developed the various equations estimated in the analysis.

RESEARCH SITES

In this study we use data from a multi-site evaluation of four batterer intervention programs (Gondolf, 1997a, 1999b). The database offers a large representative sample of batterers (n=840) across four sites and diverse regions. The modalities of the four batterer programs conform to the parameters of the prevailing state standards, which endorse cognitive-behavioral techniques taught in a group setting. The programs were selected to represent a range of format in terms of services and duration. The Pittsburgh program was 3 months in duration and provided the fewest services (referral for court-identified substance abuse or mental problems). The Denver program was the longest (9 months) and the most comprehensive program. It included individual evaluations, alcohol treatment sessions, individual psychotherapy for mental problems, and a woman's case manager. The Dallas program was 3 months with individual evaluations and additional individual counseling available, as well as women's groups available to complement the batterer group sessions. The Houston program was 5½ months with support and counseling available for the partners, as well as referrals for batterer substance abuse problems.

The programs also differed in their coordination with the courts. Most of the participants at the Denver, Dallas, and Houston programs were sent as a stipulation of their sentence following a conviction or guilty plea, or were referred from a probation officer. The Pittsburgh program differed further from the other programs in that it followed a pre-trial court order from a specialized domestic violence court. As previously mentioned, the men at the Pittsburgh program were also required to appear before the court every 30-days over a 3-month period in order to verify their compliance with the court order. A program staff member was present in the courts to document program attendance and report violations of program requirements. The other programs notified the court or probation officer about non-compliance. The response to non-compliance was less certain and much slower in this procedure. The pre-trial procedure in

the Pittsburgh court also made for shorter delays between court appearance and program intake.

The average time between court and program was 2.5 weeks in Pittsburgh, as opposed to 9 to 15 weeks at the other sites. The longer delays at the other sites were the result of court delays in fully prosecuting the case. The percentage of cases that was finally ordered to the batterer programs was about one-third of those in the Pittsburgh jurisdiction, as a result of a greater chance of the case being withdrawn or dropped during the prosecution procedures at the other 3 sites.

The police response to reassault after program intake also varied across the four sites. Approximately a quarter (21-24%) of the batterers' female partners called the police at each site sometime during the 15-month follow-up (Gondolf, 1999b). The percentage of women who reported being reassaulted during the 15-month follow-up was also comparable across the four sites (Gondolf, 1999a). Based on police records, Denver had the highest re-arrest rate with 18% of the men being arrested for domestic violence during the follow-up, whereas for Pittsburgh the re-arrest rate was 10% and in Dallas it was 3%. Houston arrest records were not available, but the re-arrest rate for domestic violence was estimated at 7% based on the women's reports.

In short, the Pittsburgh program relied on pre-trial referrals under court supervision, and the other three programs received primarily post-conviction referrals under the supervision of probation officers. Men in Pittsburgh were subjected, therefore, to swifter and more certain response to their initial arrests for domestic violence. However, the men in Denver were more likely to be re-arrested for a domestic violence offense after the program, and the men in Dallas were the least likely to be re-arrested.

SAMPLE

The first 20-25 men appearing for program intake each month were recruited into the sample, until approximately 210 men had been recruited at each of the four research sites. The refusal rate was less than 5% at each of the four research sites, suggesting a representative sampling of the program referrals for 1995.

The vast majority of the men (82%) were mandated to the programs by the courts, as opposed to others who entered the programs voluntarily (18%). The demographics of the batterers suggest a sample typical of batterers in court-referred batterer programs. The men tended to be fairly young ($M = 32$; $SD = 8.8$) and of lower socioeconomic status (64% blue-collar). Forty-five percent of the men were White, a quarter (24%) of the men were without a high school education versus a third (36%) had more than a high school education, 64% were fully employed, and 49% lived with their partners. Overall, the batterers in this sample are more racially diverse, less likely to be married, and more likely to be unemployed than batterers described in studies conducted in the Midwest (Hamberger & Hastings, 1988; Saunders, 1996) and Canada (Dutton, 1986). (See Gondolf, 1999b for additional description of the study design, sample recruitment, and sample demographics).

The subsamples for each site had similar rates of men who had previously abused or injured their partners, had been previously arrested, had parents who abused them or who had drinking problems, had alcoholic tendencies, and displayed symptoms of severe psychopathology. The variables for relationship status (i.e., marriage, children, living together) were also equivalent for the four subsamples. The site subsamples did significantly differ in terms of demographics (e.g., Pittsburgh had the highest portion of men with lower socioeconomic status and Denver, the lowest), and initial referral sources (i.e., 94% of the Pittsburgh men were court-referred, whereas 77-79% of the men at the other sites were court-referred). The post-conviction systems, as opposed to the pre-trial system of Pittsburgh, had longer time lapses between the court hearing and program intake and consequently had a

larger portion of men with new partners, and men who had no contact with their initial partners. (See Gondolf, 1999b for comparisons of subject characteristics across the four sites).

DATA COLLECTION

At program intake, trained research assistants administered a background questionnaire to the men that included questions about the incident that led to batterer program referral: a series of open-ended questions, followed by the CTS items for "physical aggression" (Straus, 1979). An alcohol screening test, personality inventory, and forms soliciting written consent to participate in the program evaluation were also administered. The background questionnaire additionally asked about the men's demographics, living situation, parent's behavior, mental health problems, alcohol use, prior treatment and counseling, abusive behavior, previous arrests, partner's response, and partner's help-seeking. On average, it took the men approximately an hour to complete the background questionnaire (15-20 minutes), the personality inventory (20-30 minutes), and the alcohol screening test (5-10 minutes).

The men's partners were interviewed by phone within 2 weeks of the man's program intake. The women were administered a background questionnaire (15-20 minutes) to verify the men's reports of abuse, alcohol abuse, mental health symptoms, previous violence, the women's response to the abuse, their help-seeking, their own alcohol and drug use, and other assistance or treatment the women may have received. Women respondents were contacted for 82% of the men ($n = 688$) at program intake. Although we were unable to contact the partners for the remaining 18% of the men, the women respondents appear to be representative of the base sample of cases. With the exception of race, there was no significant difference between the batterers' characteristics of the interviewed women and those not interviewed (female nonrespondents) in terms of demographics, relationship status, alcohol abuse, mental health problems, and criminality. This racial bias is not likely to substantially affect the research

results because the difference between minority respondents and non-respondents is relatively small, and race is controlled as a covariate in all the analyses we report.

The men and their female partners were called separately every 3 months for a 15-month follow-up period and interviewed about their relationship status, the men's behavior toward their partner, the men's alcohol and drug use, and other treatment and assistance that the men or their partners may have received. Approximately half the men reported at program intake that they were no longer living with the initial victim. They were asked if they had a new partner and how often they saw the initial partner. The initial victims also identified a small number of the new partners. Women who did not have phones or who couldn't be reached were sent a letter requesting they call a toll-free number. Respondents were paid \$10-\$20 for each completed interview.

A female partner was interviewed for 79% ($n = 662$) of the batterers at least once during the 15-month follow-up. We accounted for 67% of the women for the full 15-month follow-up. A new female partner was interviewed for 113 (14%) of the batterers, while a new partner and the initial victim were interviewed for 68 (8%) of the men. A sophisticated attrition analysis shows negligible bias resulting from attrition with regard to multivariate analyses of reassault (Jones, 1998).

In summary, the data set used for this analysis is unique in research on batterers, given its large sample size, its comprehensive set of covariates measured prior to the program, multiple sites, multiple data sources (men, women, and police records), high response rate among the batterers and their partners – especially during follow-up, and its longitudinal design with more reliable reporting resulting from follow-up interviews being conducted every three months. As discussed earlier, most previous research findings are based on small samples of less than 200 subjects, studies with low response rates of under 50%, a single follow-up interview, program completers only, single-site designs, and follow-up reports from only the

initial victim (for reviews of these studies, see Babcock, Green, & Robie, in press; Davis & Taylor, 1999; and Gondolf, 1997a).

MEASURES

In this section, we describe briefly the variables used in our study. The outcome measures and some of the most relevant predictors, such as the risk assessment instruments and batterer types, are described in detail in the body of the paper. In addition to the outcome measures, we will also describe the predictors measured at intake, program variables, predictors measured at the first 3-month follow-up, the risk assessment instruments, our coding of psychopathology, and our qualitative coding procedures. Detailed descriptions of coding for most of the predictors, however, are provided in Appendix A.

Outcome Measures

Dichotomous Outcomes

To replicate previous risk assessment research and for comparison to our multiple outcomes mode, we constructed a binary outcome reassault measure as follows. Assaults were assessed through a series of questions that included the following: (1) an open-ended question about "how was the relationship going," (2) descriptions of any conflicts and their circumstances, and (3) the inventory of physical aggression items from the CTS (Straus, 1979). An assault was considered any incident that included one of the tactics on the Physical Aggression subscale of the CTS (i.e., pushed, shoved, grabbed; slapped; hit with a fist, bit, kicked; hit with something, attempted to hit with something; choked or burned; threatened with a knife or gun; used a knife or gun; forced sex against will). The dichotomous outcome variable "any reassault" was coded 1 if any of the women (initial or new partners) reported any assault as occurring between intake and 15 months.

A second dichotomous “any reassault” variable was constructed based on women’s reports provided between the 6-month and 15-month follow-ups (which provides information on reassault that occurred between 3 and 15 months after intake). For this outcome variable, used to estimate the conditional prediction model, reassaults that occurred between intake and the first 3-month follow-up were omitted. Thus, time-varying factors measured at the first follow-up could be incorporated into the risk models without concerns about temporal order.

Multiple Outcomes Variables

The primary outcome in this study is a multiple outcomes variable, which was based on reports by the women regarding the men’s abusive behavior. Women’s reports of assault were used because of substantial underreporting by men during the follow-ups (Heckert & Gondolf, 2000b, 2000c). Men were classified in the five following categories based on their partners’ reports of abuse during the 15-month follow-up:

repeat reassaulters—more than one incident that included one of the tactics on the physical aggression subscale of the Conflict Tactics Scale;

one-time reassaulters: only one incident of physical aggression;

threatening reassault—no physical tactics but any threats (i.e., to hit, attack, or harm; to kill; to take away children or harm them, to kill or seriously harm other people; to kill or hurt himself);

controlling behavior or verbal abuse—no threats or physical tactics, but any controlling behaviors or verbal abuse (i.e., kept from talking on phone; kept from friends; stopped from going someplace; followed partner; kept from using family income; took or stole money from partner; swore or screamed; accused partner of being with another man; insulted or put down; threw, smashed, hit, or kicked something; destroyed property; or hurt a pet or pets);

no abuse—no reports of physical assault, threats, controlling behavior, or verbal abuse over 15 months.

To estimate a conditional prediction model, we also constructed a second multiple outcome variable using interviews starting at the 6-month follow-up through the 15-month follow-up (which provides information on reassault and abuse that occurred between 3 and 15 months after intake). This outcome excluded the first 3-months after intake and allowed testing of a dynamic or conditional model of risk, by allowing the conditional variables measured at the 3 month follow-up to be entered as predictors in the multivariate equations.

Missing Interviews

It is important to mention that some of the initial and new partners of the batterers in the study had missing follow-up interviews. Women were excluded from the analysis if their follow-up interviews 3, 4, and 5 (9-month, 12-month, and 15-month interviews) were missing. Thus, women's outcomes were coded if they had at least the third follow-up interview, even if the fourth and fifth interviews were missing. This decision was made to preserve as much of the sample as possible and is justified by the following. One, the majority of the reassaults occurred within the first six months; the amount of physical reassault and psychological abuse steadily declined over time. Two, we conducted an analysis with only the 348 women who had complete follow-up data for all 5 follow-ups in which we calculated their partner's multiple outcome based on the first 3 follow-ups and cross-tabulated it with the partner's multiple outcome based on all 5 follow-ups. The Kappa for this cross-classification was a robust .75, justifying the use of incomplete cases. Our decision to use cases with missing women's reports from the fourth and fifth follow-ups will result in some downward misclassification. In other words, some batterers will be classified as less severe in their cumulative measure of abuse than they actually are. For example, a small percentage of the men who are classified as one-time reassaulters over 15 months, may have actually been repeat reassaulters, if women's reports had been obtained at 12 and 15 months. We are confident, however, that the percentage of cases that are misclassified is rather small, probably less than five percent.

Predictors Measured at Intake

The predictor and control variables were taken from the background questionnaires, which were administered to the men and women at program intake. There were five main types of variables measured at intake: social background characteristics; attributes of the relationship; the batterer's previous behaviors including alcohol use, extent of contact with systems of social control, and physical and nonphysical abuse; batterer's dispositional or personality attributes, including measures of mental health; and women's perceptions of risk and previous help-seeking behavior and service contact. The variables we used have been identified as risk markers in previous research and/or have been discussed in the domestic violence literature as possibly being associated with assault. Complete descriptions of these variables are provided in Appendix A.

Social Background and Relationship Variables

The social background variables for the multivariate equations are age, race, employment status, and three indicators of socioeconomic status: occupational status, education, and whether or not the woman has received welfare or public assistance in the last 3 months. Relationship characteristics are marital status of the man at intake, living together, time involved with partner at intake, and number of children living with the man.

Previous Behaviors

Several measures of the man's alcohol and drug use and contact with systems of social control were used to assess which would be the strongest predictor of reassault. Measures of alcohol and/or drug use were the following: a binary indicator of an elevated Michigan Alcohol Screening Test (MAST) score (> 4), alcohol-related crime reported on the Michigan Alcohol Screening Test, man's report of being frequently drunk or high in the last year adjusted by the woman's report, and man's report of alcohol and/or drug treatment. Different alcohol measures

were tried because of the discrepancies between alcohol tests and self-reports. Whether the man was arrested for anything other than domestic violence in the previous year was also computed.

The women were also asked about other forms of abuse at intake, and these reports were adjusted with men's reports of the same behaviors. If either the woman or the man reported that abuse occurred, the abuse category was coded as having occurred. The women were administered an inventory of *controlling behaviors* (i.e., kept from talking on the phone; kept from spending time with friends; stopped from going some place; followed against will or knowledge; kept from using income or savings; took money from her), *verbal abuse* (i.e., swore or screamed at; accused of other relationship; put-down, insulted, called names), and *threats* (i.e., threatened to hit, attack, or harm the woman; threatened to kill any person; threatened to take away or harm children; threatened to harm other people; threatened to kill or hurt self)(see Gondolf, Heckert, & Kimmel, in press, for additional description of the non-physical abuse indices). Women's reports of the men ever committing *severe violence* prior to program intake was also included as a control variable in the prediction models, with a value of 1 indicating a "yes" response to any of the so-called severe items on the CTS (hit with a fist, bit, kicked; hit or tried to hit with something; choked or burned; threatened with or used a weapon; forced sex). In addition, the women reported whether the batterer had ever caused any other injuries other than bruises, whether they had ever sought medical help, and whether they had ever been hospitalized. These last three variables are used as indicators of severe abuse. Presumably, women who were most severely abused in the incident that brought the batterer to treatment may be at greatest risk for repeated reassault.

Psychological Characteristics and Batterer Types

Previous research has implicated personality disorders and psychological characteristics as determinants of intimate partner violence. Therefore, a series of variables pertaining to the

man's dispositional or personality characteristics were constructed using results from the Million Clinical Multiaxial Inventory, Version III (MCMI; Million, 1994). The MCMI is a 175-item test with 24 subscales that correspond to Axis II (personality disorders: e.g., antisocial or narcissistic personality) and Axis I (major disorders: e.g., major depression, schizophrenia) diagnoses. A base-rate (BR) score of 75 or above (i.e., an "elevated score") suggests that symptoms for a particular disorder are above the norm in a clinical population, or are "present". Elevated scores (BR \geq 75) on the paranoid, borderline, schizotypal, major depression, delusional, or thought disorder subscales are considered evidence of "severe" pathology. Each personality type and/or clinical disorder was coded 0=absence and 1=presence. We also explored various combinations of MCMI subscales. For example, one combined index measured elevated scores on any severe clinical syndrome (Axis I) and another measured elevated scores on any clinical personality disorder (Axis II).

Batterer types were determined through a cluster analysis of the MCMI-III data (Gondolf, under review). Four groups emerged from the cluster analysis: 1) little or no pathology, 2) narcissism or antisocial, 3) dependent/avoidant, and 4) multiple disorders. The procedures and findings are similar to those used in a cluster analysis that used the MCMI-I and a Midwest sample (Hamburger et al., 1996). The four groups in our study also approximate the prevailing batterer typologies (see Holtzworth-Munroe & Stuart, 1994), which have been presumed to be associated with reassault but have not been adequately tested as risk markers. Specifically, the little or no pathology group approximates the family violent only type proposed by Holtzworth-Munroe and Stuart. The dependent/avoidant group resembles the dysphoric/borderline type identified by Holtzworth-Munroe and Stuart. Finally, the narcissism or antisocial and the multiple disorders groups are similar to the violent/antisocial type proposed by Holtzworth-Munroe and Stuart.

Women's Characteristics

Women's perceptions of risk, previous help-seeking behavior, and service contact were obtained through the interviews conducted with women at the time of program intake. With regard to perceptions of risk, the women were asked, "How safe do you feel at this point?" (coded 1 = uncertain, not safe, in much danger, 2 = somewhat safe, and 3 = very safe), and "How likely is it that your husband will become violent towards you during the next 3-months?" (coded 1 = very likely or likely, 2 = uncertain/don't know, 3 = unlikely, 4 = very unlikely). The responses were collapsed from their original 5-point scales based on conceptual reasons and the number of responses in each category.

Women's help-seeking was measured a number of ways, including previous shelter use, prior police call, a summated help-seeking scale, and prior use of legal assistance. The woman's previous service contact was also assessed in a variety of ways, including contact by a woman's shelter, legal office, or batterers program in the past few months, alcohol or drug treatment, and other kinds of counseling or mental health treatment.

Program Variables

The effects of program context are captured by including the four sites—Pittsburgh, Denver, Houston, and Dallas—as a categorical variable. In all multivariate equations, the Pittsburgh site is the omitted reference group, because it has the greatest control through 30-day mandatory court review, the shortest duration of treatment, and the least amount of services. Thus, we will be able to ascertain how length and which additional services affect reassault outcomes, above and beyond a three-month program with few services but with mandatory court review. Court-referred (coded 0) versus voluntary status (coded 1) was determined by the referral source indicated on the background questionnaire. Court-referral included being required to attend the batterers program as a result of a criminal or civil court order or probation officer recommendation. Those who were not court-referred were attending

“voluntarily” through a referral from another social service or urging from their partner. (In Pittsburgh, 5% of the men were “voluntary” and 20% were at each of the other sites.)

Predictors Measured at First 3-Month Follow-up

Conditional variables were identified from the 3-month follow-up interval with the woman. Later follow-ups were not used for predictor variables because truly dynamic models could not be estimated for multiple outcomes, because there were not enough batterers who had repeat reassaults in any given follow-up period. Conditional variables included living arrangements, contact between partners, employment status, substance use, further arrests, and use of additional services and treatment. Time-varying predictors that were measured at both intake and follow-up 1 were employment status; living situation; child living with the batterer; use of verbal abuse, controlling behavior, threats, or physical abuse during the first three months of follow-up; arrest for non-domestic offense; the batterer had been frequently drunk or high; women’s perceptions of how likely the man was to use violence again and of safety; injuries were caused by the batterer; and medical help or hospitalization was necessary owing to injuries caused by the batterer. Conditional variables that were not measured at intake are listed below and described in detail in Appendix A.

Program dropout was derived from the number of sessions the participants attended at each program according to program records. This information was converted into a dichotomous variable, coded 1 if the batterer attended at least 12 sessions and 0 if he did not. We used the 12-week threshold to accommodate the minimum required sessions of 12 weeks at two of the research sites (Pittsburgh and Dallas) and what appears to be a threshold regardless of program length (i.e., more than 90% of the dropouts in the longer programs occur within the first three months). Reasons for dropping out vary. Some men were formally dismissed for reassaulting, whereas others simply withdrew voluntarily from the program.

Relationship and Behavioral Variables

To reflect the dynamic effects of relationship and behavioral contingencies, a number of conditional variables were used that measured the quality of the batterer-partner relationship, as well as other behaviors that might possibly be related to spousal assault. Relationship variables used in the analysis included men's and women's reports of communication problems, conflicts, and poor sexual relations. Only the men were asked if there was anything that their partner had done that angered them or might have "set them off" during the previous three months and if they had any close calls in which they felt like hitting their partners. The "anger" and "close calls" variables are indicators of conflict in the relationship that may not be tapped by the other predictors. They may also tap into behavioral intentions or a predisposition to use violence, to which the men may not be willing to directly admit. Additional relationship and behavioral variables that occurred in the previous three months based on the women's reports were emotional injury, hospital visit by the woman, need for a medical visit (but not taken), and had the man been abusive to any children.

Women's Help-seeking

A large number of help-seeking variables were measured at the first follow-up to reflect what happened over the previous three months including shelter contact, legal action, police call, protection order, divorce filing, woman received various types of counseling or alcohol and drug treatment, woman received any kinds of assistance or social services, and use of informal services by the woman. A total quality of life indicator was also computed.

Effects of Intervention

A number of conditional variables were used to determine whether there were possible effects of batterer intervention. Binary indicators were constructed to reflect the men's and women's perceptions of what the batterer had done to avoid incidents of violence during the

past three months, including "time outs," avoidance strategies by the batterer, and discussion oriented strategies. A variety of indicators of changes in the man were also examined, along with indicators of suggested changes in the batterer programs.

One potentially beneficial effect of program participation is the possible deterrent effect of making batterers aware of the potential sanctions for their antisocial behavior. Perceptions of sanctions were assessed with a measure of severity and a measure of certainty for batterer dropout and reassault. At the first 3-month follow-up, the batterers were asked two open-ended questions with regard to what would happen if they "are physically abusive again" and "if you stop attending the program or have stopped." Interviewers checked all applicable responses, which we grouped into four types of perceived sanctions: none (no sanctions or uncertain), extralegal sanctions (partner would leave, counselor/group would talk to man, other), preliminary legal steps (partner would call police, man would be arrested, sent back to court, dismissed from program), and legal sanctions (fined, sent back to program, jail). We present all that are mentioned as well as the most severe mentioned. This variable serves as our measure of severity of sanctions (see Heckert & Gondolf, 2000, for additional description of these perceptions measures).

In addition, the batterers were asked how likely it was "that you'd be put in jail" for program dropout and partner reassault. The five-level Likert-type response was collapsed into (1) likely, (2) don't know/uncertain and other, and (3) unlikely categories. This variable serves as our measure of certainty of sanctions.

Other Interventions

Both the men and the women were asked if the man received any other assistance over the previous three months, such as alcohol or drug abuse treatment, individual counseling, other types of counseling other than the batterer program and other forms of assistance like church attendance, recreational activity, parenting program, special clubs or support groups, job

training or placement program, reading self-help books or materials, medical help, or any other type of support.

Level of Contact

One additional factor measured after intake at each follow-up interview was amount of contact. Two variables were constructed from this question. First, men who had no contact with their primary partner between the 3-month and 15-month follow-ups were coded 1, whereas men who had contact at any time during the 3- to 15-month follow-up period were coded 0. Second, men who had no contact during any of the 3-month intervals were coded 1, whereas men who had contact during each of the 3-month intervals were coded 0. Although no-contact cases have a slightly lower risk of reassault, reassaults still occur because the batterer may pursue a woman who leaves. As Gondolf (1997b) has suggested, no contact appears difficult for women to achieve, because of issues such as child visitation, financial dependence, and pursuit by some men.

Simulated Risk Assessment Instruments

In this section, we describe briefly the three risk instruments that we were able to simulate using our data: the Kingston Screening Instrument for Domestic Violence Offenders (K-SID), the SARA (Spousal Assault Risk Assessment), and Campbell's Danger Assessment Scale (DAS). The DAS was designed to assess the potential for lethality, whereas the K-SID and the SARA were designed to assess the risk for further violence and not necessarily lethality. Tables 1, 2, and 3 show the items on each instrument and their availability in our data set. These are three of the most popular risk instruments currently in use, either formally or informally.

The K-SID was developed by Richard Gelles and is used statewide in Connecticut for sentencing in domestic violence cases (Roehl & Guertin, 2000). It has three parts: the poverty chart, a severity and injury index, and an index of ten risk markers. Total K-SID scores are

classified into risk categories, although we assess how well both the total scores and the risk classifications predict reassault. A study of the K-SID is currently underway by Eleanor Lyons in Connecticut, but prediction results are not available at this time. As can be seen in Table 1, our simulated version of the K-SID is excellent. The only item that is completely missing out of the eleven items is previous violation of a protection order. We have seven of the eleven items available, and similar items for the remaining three items.

The DAS was developed by Jacqueline Campbell (1995) based on known risk markers for lethal violence. The scale consists of fifteen yes or no questions; scoring is based on the total number of yes answers. There are no cutoff scores, risk categories, or weighting of the items. Five studies have shown very good reliability and moderate construct validity (Roehl & Guertin, 2000). A study by Goodman, Dutton, and Bennett (2000) showed that the DAS was modestly predictive of misdemeanor reassault in the short term. However, their original sample was quite small ($n = 92$) and their follow-up sample was even smaller ($n = 47$), which limits the utility of their findings. A recent study by Weisz, Tolman, and Saunders (2000) also simulated the DAS and found it to be a modest predictor of reassault at a 4-month follow-up when used in conjunction with other significant predictors. At this time, the DAS is not being formally used in sentencing (Roehl & Guertin, 2000). As Table 2 shows, our simulated version of Campbell's Danger Assessment Scale is very good. Three of the fifteen items in the index are completely missing, five of the items are somewhat to very similar, and the remaining seven items are available in our data set.

The SARA was developed at the British Columbia Institute on Family Violence (Kropp, Hart, Webster, & Eaves, 1999). It consists of a clinical checklist of twenty risk factors, which are grouped into five areas: criminal history, psychosocial adjustment, spousal abuse history, current offense characteristics, and other (Roehl & Guertin, 2000). Each of the twenty items is scored 0 to 2; risk is considered to increase with the number of items that are present (score of 2). We use total scores in this study, number of factors present, and risk categories, to see

which scoring method is most predictive of reassault. The SARA is currently used by courts in two states (Colorado and Vermont) to determine level of supervision and intervention, and in Canada to determine treatment plan based on level of risk (Roehl & Guertin, 2000). It relies on victim and offender interviews, as well as criminal records. Although interrater reliability is high and internal consistency is moderate, evidence of predictive validity is modest. Only enhanced clinical judgments predict recidivism, and the prediction is modest (Kropp et al., 1999; Kropp & Hart, 2000). A large-scale study is being conducted in Colorado to assess the predictive validity of the SARA. Preliminary results suggest that the SARA predicts reoffense better than chance (Dutton & Kropp, 2000). As Table 3 reveals, the simulated version of the Spousal Assault Risk Assessment instrument is very good as well. Four of the twenty items in the index are completely missing, six of the items are similar, and the remaining ten items are available in our data set.

It is important to note that the exact instruments were not administered in this study. The risk instruments can only be simulated by our data. However, we believe there is utility in the simulation since the implementation of these instruments in actual practice frequently results in missing data. For example, a recent study using the SARA was able to interview only about thirty percent of the victims (Houghton, 1999). It is arguable that our dataset has information that is more complete than many "real world" implementations of the actual risk assessment instruments. We suggest, therefore, that our "simulated" versions of the three risk assessment instruments are reasonable proxies for the actual instruments (i.e., they simulate actual use of the instruments).

Coding of Psychopathology Using the MCMI-III

For our exploration of alternate subtypes of batterers, we developed measures of primary psychopathy, secondary psychopathy, and psychopathic tendencies using MCMI-III profiles (Gondolf & White, 2001). We followed the interpretations recommended in the

instrument manuals with one revision. We identified psychopathic tendencies according to profile configurations recommended by experts on psychopathy and the MCMI (Blackburn, 1998; Millon & Davis, 1998). Any evidence of psychopathic tendencies was given priority over other possible interpretations in order to ensure the maximum inclusion of such tendencies. The broader and more liberal conception of psychopathy is likely to identify more men than narrower conceptions previously used in the field. We will describe our procedures in more detail in the results section.

Qualitative Coding of Violence Narratives

We conducted qualitative case reviews of women's narratives and atypical cases to determine whether there are behaviors or situations, rather than personality types, which distinguish the repeat reassaulters. To explore for alternative abuse outcomes, our research team coded the women's descriptions of the violent incidents using a sequential, situational conception of violence (Monahan, 1996; Mulvey & Lidz, 1993). These descriptions of violence were prompted by an open-ended question that preceded the administration of the CTS, which asked the women to tell the interviewer about what happened in the most severe incident of physical abuse. The interviewers then coded the issues, circumstances, precipitants, alcohol use, man's emotional state, couple interaction, pattern of tactics, and the woman's and the man's response to the violence. The codes for the various components were, then, cross-tabulated with the categories for reassault (no, once, repeat) to identify differences across the outcomes. The assistants also wrote their overall impressions and observations of the violence in each case, and other researchers summarized this information and used it to confirm and elaborate the cross-tabulations.

In an additional attempt to identify distinguishing characteristics of repeat reassaulters, we identified "atypical" men and women in our evaluation. Our rich database and extensive interviewing enabled us to do this. At weekly meetings of the research assistants conducting

the interviews, the interviewers nominated any atypical cases among their interviewees. The "atypical cases" were considered any man, woman, or couple who were distinct, unique, or exceptional in terms of their characteristics or circumstances. The research assistants and supervising researchers then discussed the nominations, compared them to other cases and previous nominations, and agreed to classify the cases as "atypical" or not. At a later time, two research assistants, who were not involved in the interviewing, read through the fifteen months worth of interview records for each atypical case and wrote a case summary. The case summaries were then grouped according to the distinguishing features of the cases and will be discussed in the results section. Brief examples of atypical cases include cases of severe child abuse and molestation ($n = 11$), severe violence from the woman ($n = 4$), and resounding "successes" ($n = 11$).

GENERAL ANALYTIC PROCEDURES

In this section, we review the general analytical procedures used to conduct the quantitative analyses. We discuss multinomial logistic regression and the procedures we used to build the various models discussed in the results. Multinomial logistic regression was used to accommodate the multiple outcomes that were an essential feature of the analysis. Specific details of some of the analyses will be provided in the results sections for ease of understanding.

Multinomial Logistic Regression

We use multinomial (polytomous) logistic regression to test our intake (static) and conditional (dynamic) prediction models of multiple outcomes and to assess the efficacy of risk assessment instruments and batterer typologies in predicting repeat reassault among batterers who have been referred to batterer treatment. These multinomial models are compared to binary logistic regression equations that model a dichotomous outcome variable (no reassault

versus reassault), but otherwise are identical with regard to predictor variables. The comparison of the correct classification in the two procedures helps us to identify a possible improvement in prediction.

There are different types of statistical procedures used to analyze dependent variables with multiple outcomes. Which procedure to use depends on the level of measurement of the outcome variable (nominal, ordinal, or interval) and whether various assumptions are met by the data. We conceptualize the outcome categories as constituting an ordinal dependent variable, ranging from non-abuse to repeat reassault. It is possible, however, that the outcomes actually operate as discrete, nominal outcomes. Thus, we need to use a statistical procedure that has less restrictive assumptions, can accommodate nominal and ordinal level variables, and which can help us determine how the outcome variable actually operates vis-à-vis a set of predictor variables (as either nominal or ordinal).

The appropriate analytical technique to use with an ordered multiple outcome variable is either logistic regression or probit analysis (DeMaris, 1995; Hosmer & Lemeshow, 1989). When the outcome is ordinal, ordinary least squares regression produces biased estimates of the regression coefficients for the independent variables, since it assumes an interval level dependent variable. The use of ordinary least squares is particularly problematic because of the assumption of a linear function, the assumption that the error term and the predictors are uncorrelated, and the assumption that the variance of the errors is constant across combinations of the predictor values (error homoscedasticity)(DeMaris, 1995). Both probit analysis and multinomial logistic regression correct for the errors in variance produced by the use of ordinary least squares and also take into account the multiplicative and nonlinear nature of the effects of the independent variables (Chang & Tuckman, 1989). The use of probit analysis versus logistic regression is determined by the assumption that is made about the distribution of the error term (DeMaris, 1995). Probit is used if one assumes a normal distribution of the error term, and logistic regression is used if one assumes a logistic

distribution. Because these distributions are sufficiently similar in shape, the choice of distribution is not of much practical consequence; substantive conclusions will be quite similar (DeMaris, 1995).

The primary advantage of multinomial logit analysis is that it is more practical and feasible than probit if the dependent variable is not ordinal. Multinomial probit is logically possible, but impractical, with an unordered nominal dependent variable (Aldrich & Nelson, 1984). Discriminant analysis can also be used if the outcome variable is nominal; however, it requires restrictive assumptions that may not be met by our data. For example, discriminant analysis requires the assumptions of multivariate normality among the independent variables and that the variance-covariance matrices within each group are equal (Klecka, 1990). Consequently, we used multinomial logistic regression because the ordinal nature of the dependent variable is arguable, and because of its greater flexibility, ease of interpretation, and increasing usage in the social science literature in the past decade.

Multinomial logistic regression, moreover, shows the direction of predictors' effects on the outcome variable (DeMaris, 1995). One can determine if the predictor variables produce monotonic shifts in the odds of being more abusive versus less abusive over the 15-month follow-up period. If monotonic shifts occur, the assumption that the outcome variable should be treated as ordinal is supported. To test whether the outcome variable appears to be ordinal, the score test for the proportional odds (or parallel lines) assumption will be examined. This test allows us to determine whether the effects of the predictor variables can be summarized with one equation (ordered or cumulative logistic regression), or whether multiple equations are necessary (unordered or nominal logistic regression) to model the log odds of being less versus more abusive over the 15-month follow-up period (DeMaris, 1995).

The use of multinomial logistic regression is still relatively uncommon in the social sciences. The greater availability of appropriate computer software and a better understanding of this technique make its usage more feasible and will no doubt provide for more appropriate

prediction models in the future. Its usage could constitute a significant advance over previous prediction analyses that have only modeled a dichotomous outcome variable.

Building the Models

To test the intake (static) prediction model, the multiple outcome variable was regressed on the variables measured at intake. To model a multiple outcome variable with five categories, four equations that estimate log odds are computed. We use "repeat reassault" as the baseline category and determine how the predictor variables affect the log odds of being in each of the remaining four categories. The repeat reassault group is used as the baseline category because they are the group of most interest and are responsible for most of the severe and injurious assaults. The model chi-square is a global test of the significance of the overall model. With this test we can determine if the intake variables distinguish between repeat reassaulters and each of the other types of outcomes. There is also a global test available in SPSS (Nominal Logistic Regression Procedure) to determine which predictor variables have significant effects overall in distinguishing categories of program outcomes.

Nature of the Multiple Outcome Variable

It is important to decide whether to treat the outcome variable as ordinal and present results from an ordered (proportional odds or cumulative) logistic regression, or whether to treat the outcome variable as nominal and present results from an unordered (nominal) logistic regression. Although the multiple outcome variable is conceptually ordinal, the types of abuses represented may function as discrete categories (i.e., unordered, nominal). Therefore, we ran the multinomial equations both ways – with ordered logistic regression and nominal logistic regression. There are essentially two ways to determine which analysis is most appropriate. First, a global test of the proportional odds or parallel lines assumption (available in SPSS) assesses whether the dependent variable functions as an ordinal or a nominal level variable by

evaluating whether the effects of the predictor variables are monotonic or non-monotonic. If the effects are monotonic, this provides evidence that the outcome variable operates like an ordinal variable. Otherwise, the outcome variable should be treated as if it were nominal.

There is a second procedure to assist in deciding whether to treat the outcome variable as nominal or ordinal. The relative goodness-of-fit of the ordered logistic regression can be assessed by comparing the likelihood value with that obtained by the nominal logistic regression. The "likelihood ratio" test basically compares -2 times the differences in the log-likelihood values for the two models to the appropriate chi-square distribution. A large difference in the log-likelihoods of the two models suggests that the nominal logistic regression provides a better fit to the data (StataCorp, 1999). If the nominal logistic regression model fits the data better, odds ratios will be calculated to determine the substantive effect of each predictor variable. SPSS also provides a classification table for nominal logistic regression that allowed us to ascertain what percentage of the repeat reassaulters were predicted by the equation to be repeaters, the false positive rate, and the overall percentage of batterers who were correctly predicted with regard to their abuse status.

Intake Models

To build the multiple outcomes models, we proceeded hierarchically, testing one variable at a time within blocks. The multinomial procedures for both SPSS and STATA do not have automatic stepwise procedures. We first entered referral status, age, and race as control variables that were kept in the models regardless of whether they were significant predictors. We then entered social background variables one at a time and retained variables that had at least one significant coefficient at $p < .10$ (two-tailed) in any of the four equations. We also retained the variable if the overall global test for the predictor was significant. In addition, we examined the classification tables available in SPSS and retained variables that significantly improved overall classification or classification of repeat reassault. The following blocks of

variables were entered sequentially: sociodemographics, socioeconomic indicators, relationship variables, previous abuse and other related behaviors, personality and mental health variables, women's perceptions and help-seeking variables, and program site. The variables were entered in that order based on how easy it is for clinicians and criminal justice personnel to collect such data for risk assessment, as well as based on their logical order in an implicit causal model. Thus, only variables that explained unique variance in the outcome variable were entered after the initial block of predictor variables was entered. Once the full model was built, we then trimmed the models in a backwards fashion by removing variables, one at a time, based on the global test of the null hypothesis that the predictor variable had no effect on the dependent variable. With the exceptions of referral status, age, race, and program site, variables were removed if their global p-values were greater than .20 (two-tailed), and they did not contribute substantially to accurate classification of repeat reassaulters and overall classification.

Conditional Models

The conditional prediction models were built in the same fashion. Static factors measured at intake were entered first (referral status, age, race, and program site in the first block; education and occupational status in the second block; marital status and time in relationship in the third block; and personality and psychological variables in the fourth block), followed by dynamic factors measured at the first follow-up. The conditional factors were entered sequentially in the following blocks (though one at a time): program dropout in the fifth block; men's employment status in the sixth block; access to the woman variables (living together, no contact) in the seventh block; relationship variables in the eighth block; abusive and other antisocial behaviors in the ninth block; women's perceptions in the tenth block; and women's help seeking and men's changes in the last block. Once again, the sequential order of the blocks was based on how easy it would be for the criminal justice system to collect the

information and their temporal priority in an implicit causal model. The conditional models were compared to the intake models based on their ability to accurately classify men into the five outcomes (i.e., according to the correct classification tables).

PART V: RESULTS AND DISCUSSION FOR MULTIPLE-OUTCOME ANALYSES

INTRODUCTION

To accomplish our primary objectives, we developed a conditional prediction model of multiple outcomes. First we used multinomial logistic regression to compute an intake model of multiple outcomes, using risk markers measured at intake as predictors. We compared this model to a binary logistic regression model that regressed a dichotomous outcome (any reassault versus no reassault) on the same intake risk markers. We did this to ascertain whether previous prediction efforts may have been limited by artificial dichotomization of program outcomes into "success" and "failure." Second, based on research in the psychiatric violence literature, we explored whether consideration of conditional risk markers would improve prediction over intake models. To do so, we used multinomial logistic regression to build a conditional prediction model by using intake risk markers and risk markers measured at the first 3-month follow-up to predict multiple outcomes that occurred between the 3-month and 15-month follow-ups. Again, we compared this multinomial model to a dichotomous model using the same predictors to assess the utility of multiple outcomes.

Third, we investigated whether prediction could be improved by using risk instruments. We assessed the predictive power of simulated versions of three popular risk instruments and compared it to the predictive power of the intake and conditional models, as well as to the predictive power of women's perceptions considered alone. Fourth, considerable effort by researchers has been devoted to developing meaningful "batterer types," in hopes that they would improve prediction of reassault. To test this conjecture, we entered a measure of batterer type (that parallels standard types in the literature) as a risk marker in the multinomial equations. We then entered an interaction term of "batterer type" by program format to assess whether different types of batterers might benefit from specific program formats.

RISK MARKERS FOR INTAKE MODELS

Dichotomous Intake Model

To determine whether the multiple outcome measure would improve prediction, we first estimated logistic regression models using a dichotomous outcome of "any reassault" versus "no reassault." Either the initial partner or a new partner reported that around 33% of the men committed at least one reassault over the 15-month follow-up. As Table 4 reveals, the binary logistic regression results confirmed previous research (see Dutton et al., 1997; Gondolf, 1997b; Hamberger & Hastings, 1990; Hanson & Wallace-Capretta, 2000; Saunders, 1995; Tolman & Bennett, 1990). Significant predictors that increased the likelihood of reassault included younger age, living with partner, having no children living with the man, heavy drinking, use of emotional abuse and threats, prior "non-domestic violence" arrest, the woman's report of likely violence in the next three months, low help-seeking by the woman, and the woman's use of shelter or other social services. White men were significantly more likely to reassault than Latino men in this sample. Overall, the dichotomous model was significant ($X^2=119.43$, $df=23$; $p<.001$) but modestly, at best, predicted reassault above and beyond chance (overall accuracy = 75%; sensitivity = 44%). The sensitivity rate is especially disappointing with regards to risk assessment; only 44% of the men who subsequently reassaulted were predicted to do so by the equation. The rate of false positives (men who were predicted to reassault, but who did not) was only 11%. This rate, in conjunction with the low sensitivity rate, reveals that the model tends to over classify men as not likely to reassault, producing a high percentage of false negatives. To be clinically useful, the rate of false negatives needs to be reduced. Otherwise, men who are at high risk for reoffending may not be identified for the intensive intervention they warrant.

Multiple Outcomes Intake Model

For comparison to the binary logistic regression, we conducted a multinomial logistic regression analysis, again using only variables collected at program intake. The analysis was based on approximately 499 cases for which the multiple outcome variables could be constructed and data were available on the predictor variables. The distribution across the five outcome groups for these 499 cases is: no physical or psychological abuse (19%); controlling behavior/verbal abuse only (26%); threats with no physical reassault (20%); one-time reassault (12%); and repeat reassault (23%). Therefore, about two-thirds of the men who reassaulted were repeat reassaulters.

As discussed in the procedures section, we developed two conditional, multinomial logistic regression equations, using the intake predictors described in the methods section: an ordered multinomial logistic regression (cumulative log model or proportional odds model) and an unordered multinomial logistic regression for comparison. Based on a likelihood ratio test and tests of the assumptions of parallel lines, the results demonstrated that the unordered multinomial model was significantly better than the ordered model. The difference in the chi-squares for the two models was substantial (unordered model $X^2 = 270.61$, $df = 108$; ordered model $X^2 = 128.90$, $df = 27$), and the "likelihood ratio" test suggested the models were significantly different ($X^2 = 171.46$, $df = 81$, $p < .01$). Further, the proportional odds or parallel lines assumption necessary for the ordered model was not met ($p < .001$); the null hypothesis that the slope coefficients were equal across response categories was strongly rejected ($X^2 = 171.46$, $df = 81$). The multiple outcomes variable should, therefore, be treated as a nominal outcome variable, rather than an ordinal outcome variable. The multiple outcome categories do not necessarily represent a progression of least to most severe abuse.

We examined the multinomial logistic equations to assess which outcome categories were best predicted or distinguished and drew a number of conclusions. First, the sets of predictors that distinguished the "repeat reassault" from the "no abuse" category and the "repeat

reassault" from the "controlling/verbal abuse" category are quite similar (see Table 5 for actual coefficients; see Table 6 for classification results). Thus, "no abuse" and "controlling/verbally abusive" outcomes are difficult to distinguish based on variables available at program intake. Second, the best discrimination by variables available at program intake is between "repeat reassault" and the "no abuse" and "controlling/verbal abuse" categories. There are fewer variables that discriminate between the "repeat reassault" and "threats" categories. Third, although there are fewer factors that discriminate between the "repeat reassault" and the "one-time reassault" categories, the odds ratios suggest they are reasonably strong predictors. Fourth, as Table 6 indicates, the overall model is moderately successful at predicting repeat reassaulters (sensitivity = 70%), and the rate of false positives (men who are predicted to be repeat reassaulters who are not) was relatively low (16.3%). However, the overall percentage of batterers who were accurately predicted based on intake variables was only 48%. Nonetheless, prediction of "repeat reassault" in the multinomial equation is superior to prediction of "any reassault" in the binary equation, with a 26% improvement in sensitivity.

Comparison of Multiple Outcomes and Dichotomous Outcome Intake Models

In comparing the results from the multiple outcomes model to the model with dichotomous outcomes, several points emerge. One, the intake variables predict "repeat reassault" with much greater sensitivity than "any reassault." Two, overall prediction is stronger with the dichotomous model, primarily because the intake variables do not distinguish very well between the "no abuse" and the "controlling/verbal abuse" categories in the multinomial outcome equation. Based on these results, future research should attempt prediction with fewer outcomes, perhaps by starting with: 1) no abuse or emotional abuse only; 2) threats with no physical reassault; 3) one-time reassault; and 4) repeat reassault, which should improve overall prediction, although perhaps not prediction of repeat reassault. Regardless, consideration of multiple outcomes is important for improving prediction of the highest risk batterers.

Risk Markers for Repeat Reassault

Table 5 summarizes the risk markers that distinguished the repeat reassaulters from each of the other categories. With the exception of women's perceptions, we discuss the most consistent predictors first in the order they were entered into the model. We discuss women's perceptions in a separate subsection, because they are not, strictly speaking, objective, actuarial risk markers. Instead, they are risk assessments or judgments made by the batterers' partners, much like clinical judgments.

In general, age is a very consistent predictor across the four equations; men in the youngest age group (18-26) are more likely to repeatedly reassault. Race is also a consistent predictor. Consonant with the dichotomous results, Whites are more likely to repeatedly reassault than African American men and Latino men in this sample. Living together at intake, having children living with the batterer, and the woman receiving injuries are consistent risk markers for repeated reassault, although they do not distinguish one-time reassaulters from the repeat reassaulters. Living together at intake and the woman receiving injuries increase the likelihood of repeat reassault, whereas having a child(ren) living with the batterer decreases the likelihood of repeat reassault. Previous psychological abuse in the form of either verbal abuse or controlling behavior was an important predictor of repeat reassault in all the equations, as was the woman using shelter or social services.

Several variables were less consistent in their prediction. Working as a laborer or a service worker decreased the likelihood of repeat reassault relative to verbal abuse/controlling behavior and threats, but was not significant in the other equations. Likewise, having a non-domestic violence arrest prior to intake increased the risk of repeat reassault relative to emotional abuse but was not a significant risk marker in the other equations.

Women's Characteristics and Perceptions

Because women's characteristics and perceptions are not typically included in prediction research, we assessed how much they contribute to prediction of reassault above and beyond men's characteristics and reports. As Table 7 reveals, when only men's variables and reports are considered, the sensitivity rate of the multinomial outcomes model is only 55%, with an overall classification rate of 38%. Inclusion of women's characteristics other than women's perceptions increased the sensitivity rate slightly to 58% (overall classification = 42%). Next, inclusion of women's perceptions of risk increased sensitivity by 12% to the sensitivity rate of 70% in the final model. Women's perceptions of risk, then, substantially improved prediction in the multiple outcomes model over and above other risk markers.

As Table 5 shows, women's perceptions consistently predicted repeat reassault across the four equations, but their relationship to repeat reassault is not straightforward. In general (in 3 of the 4 equations), if the woman felt somewhat safe, the man was more likely to repeatedly reassault than if the woman felt very safe. However, if the woman did not know, felt not safe, or felt in much danger, the man was not significantly more likely to repeatedly reassault than if the woman felt very safe. Upon reflection, this finding makes sense. When women feel greater risk, they are more likely to respond in ways to reduce that risk, either by leaving the batterer, engaging in safety planning, and the like. If women respond that they feel somewhat safe, they may be having some uneasiness or uncertainty about the risk of reassault, but not enough to take proactive action to reduce the actual risk. This finding is of importance to advocates who work with women. If the women do not assert that they feel very safe, it would be reasonable to inform them to trust their instincts if they feel even slight uncertainty. They should be counseled to pay close attention to any warning signs of impending assault.

In two of the four equations, if the woman was uncertain or did not know if the man would use violence in the next three months, the man was more likely to be a repeat reassaulter over 15 months. If the woman perceived violence as unlikely, then the man was less likely to be

a repeat reassaulter than if they viewed violence as very unlikely. Again, women's advocates should work carefully with women to ascertain their perceptions. If women are uncertain that violence will be used, they should be cautioned that violence is a distinct possibility. However, it appears that a subset of women may put themselves at greater risk because they perceive violence as very unlikely and therefore do not take necessary precautions because of this possible denial. This finding may also reflect the very dynamic nature of some of these relationships. Future research needs to explore in more depth the precise nature of the relationship between women's perceptions and reassault, as well as the impact of women's perceptions on their safety-related actions. Dynamic models should also be developed to see how well women's perceptions predict reassault in the short-term (e.g., over 3 months), as opposed to the models that we developed that predicted reassault over 15 months.

Conclusions and Implications

A number of conclusions and implications follow from this analysis. First, the results clearly suggest that researchers and program evaluations need to distinguish between one-time reassaulters and repeat reassaulters. It is especially important to make this distinction when conducting risk assessment. Prediction of "repeat reassault" was superior to prediction of "any reassault" and identified risk markers that were not present in the dichotomous equation. Developing risk assessment instruments that predict repeat reassault or serious offending may be, therefore, more fruitful and productive than risk instruments that have been developed to predict one-time reassault or reoffending.

Second, previous emphasis on personality traits and batterer types may be misdirected. Once other important predictors were controlled, none of the personality, mental health variables, or various classifications of batterer types were significant predictors of reassault or repeat reassault. This point does not deny the possibility that the effects of personality and psychological characteristics might be indirect, as mediated by conditional factors, (or

moderating) in some underlying causal model; however, it demonstrates that such measures are *not necessary* (or useful) for accurate risk prediction.

Third, women's perceptions of risk were consistent predictors of reassault, especially repeat reassault. The inclusion of women's perceptions of risk will likely enhance risk assessment efforts. The women may be able to assess risk in broader and more idiosyncratic ways than a combination of risk markers.

Fourth, variations in program format did not influence the type of outcome, which reinforces previous findings reported from this study, which examined dichotomous outcomes (Gondolf, 1999b; Heckert & Gondolf, 2000a). Fifth, although prediction is improved using multiple outcomes, the level of prediction still has questionable clinical utility. The thirty percent rate of false negatives is still quite high and suggests that risk markers should be used with great caution by clinicians and practitioners.

In sum, the first hypothesis was partially supported. Prediction of reassault was improved with a multiple outcome, but it is still relatively modest overall.

RISK MARKERS FOR CONDITIONAL PREDICTION MODELS

The second stage of our risk marker analysis was to test a conditional model of prediction. We entered conditional variables measured at the 3-month follow-up into the logistic regression equations, using the multiple outcome variable based on violence that occurred between 3 and 15 months as the dependent variable. There is an emerging consensus that using conditional or dynamic predictors will improve prediction.

Although conditional factors emerged as significant risk markers in our multinomial equations, the overall prediction, based on accuracy of classification, was not as strong as overall prediction for the intake models described above (see Table 8). However, it is important to note that the distribution across the multiple outcomes changed as a result of eliminating the assaults that occurred between intake and the first 3-month follow-up. As Table 9 reveals, only

15% of the men were repeat reassaulters between 3 months to 15 months, and 12% were one-time reassaulters during that same time period. This restriction in the variability of the outcome measure makes it harder to obtain accurate prediction of the repeat reassaulters. Additionally, the analysis is based on fewer cases (n=399) than the intake model (n=499) owing to listwise deletion of missing values (and the inclusion of conditional variables measured at follow-up one, which had more missing values). Thus, it is difficult to compare the results with the static intake model described above. It is probably more appropriate to assess whether conditional variables emerged as risk markers, and whether the conditional variables increased sensitivity and overall predictability above and beyond the static risk markers in the models discussed below.

Based on that logic, it is safe to conclude that incorporating conditional variables improves prediction, although not dramatically. As Table 8 indicates, conditional variables do emerge as important predictors of repeat reassault. In fact, the three most consistent predictors of repeat reassault in the conditional model were conditional variables. Specifically, a physical assault between intake and the first follow-up, the woman filing for a protection order between intake and the first follow-up, and the woman reporting negative changes in the man occurring after intake all increased the likelihood of "repeat reassault." Other conditional variables also contributed to prediction. Batterers whose partners report that they are having any troubles in their relationship were more likely to "repeatedly reassault," and this variable was one of only three risk markers that distinguished "one-time reassault" from "repeat reassault." Heavy drinking between intake and the first follow-up significantly distinguished between the "no abuse" and "repeat reassault" categories. If the man used any verbal abuse between intake and three months, they were more likely to "repeatedly reassault" than to be "non-abusive." If the man received drug treatment between intake and the first follow-up they were more likely to "repeatedly reassault" than to exclusively engage in "controlling/verbally abusive" behavior.

Several static risk markers were also predictive in the conditional models. Age was a variable that distinguished the "threats only" group from the "repeat reassault" group, with

younger men being more likely to "repeatedly reassault" than men over the age of 35. Program site was another important static predictor of "repeat reassault" between three and fifteen months. In general, batterers in Dallas were more likely to "repeatedly reassault."

Despite the weakened prediction relative to the earlier models, the conditional model is relatively strong. Conditional variables are consistent predictors and many have moderate to strong odds ratios. Further, the Cox and Snell pseudo R-square is a very robust 53% ($X^2=301.72$; $p < .001$). An important reason for the weakened prediction is the inability of the model to distinguish between one-time reassaulters and repeat reassaulters. Only two static and three conditional variables distinguished between the two reassault groups, which reduced the sensitivity rate for repeat reassault (see Table 10). With conditional prediction, repeat reassault may not be the best way to categorize high-risk groups, because the distinctions between one-time reassaulters and repeat reassaulters emerge over time. It might be best to categorize high risk based on severe versus minor reassaults, or based on lethality or another measure of "dangerousness." Future research should, therefore, develop conditional models with a different set of multiple outcomes, perhaps: 1) no abuse or controlling and verbal abuse only; 2) threats with or without controlling and verbal abuse; 3) minor reassault; 4) severe and/or repeated reassault. Further collapsing outcome categories (e.g., categories 1 and 2) may possibly improve prediction as well.

Although conditional variables emerged as important predictors of repeat reassault, the conditional model did not substantially improve prediction over the static model. Thus, the second hypothesis regarding prediction with a conditional model and multiple outcomes was not supported. The argument that the use of dynamic risk prediction will markedly improve the prediction of abuse and reassault is not supported by our results. We do acknowledge, however, that our (3 to 15 month) conditional model is not a true dynamic model in the sense of predicting short-term risk, as can be done with Generalized Estimating Equations (GEE) models or proportional hazards models (see Jones & Gondolf, 2001). The conditional variables were

based on only the intake to 3-month follow-up interval, rather than for each of the five follow-up intervals (because the incidence of repeat reassault was too low in each follow-up period).

PREDICTION BY RISK ASSESSMENT INSTRUMENTS

As outlined in the objectives section, the second area we tested was the effectiveness of simulated risk assessment instruments in predicting abuse and reassault outcomes. Three popular risk instruments were simulated with our data – the Kingston Screening Instrument for Domestic Violence Offenders (K-SID), the Spousal Assault Risk Assessment instrument (SARA), and Campbell's Danger Assessment instrument (DAS). As Table 11 shows, the K-SID scores, whether used as a total score or classified into low, medium, and high risk, by themselves gave very weak prediction of multiple outcomes. The ability of the simulated K-SID to accurately predict repeat reassault was extremely low (11% sensitivity rate). The simulated SARA total scores predicted significantly better than the K-SID scores, but still offered modest prediction of multiple outcomes. The sensitivity rate for repeat reassault using the SARA total scores was a modest 43%. Using a high risk cut point for the SARA scores substantially reduced prediction from the SARA total scores, dropping the sensitivity rate to 27%.

The simulated risk assessment instrument that predicted best was the DAS. In fact, simulated DAS scores based on women's reports accurately classified a reasonably robust 66% of the repeat reassault, although the rate of false positives was 33%. In other words, 33% of the men who did not repeatedly reassault were predicted by the DAS to be repeaters. Examination of the logistic coefficients for the SARA scores and the DAS scores (not shown) reveals that the simulated SARA does poorly at distinguishing the "threats only" outcome and "one-time reassault" from the "repeat reassault" outcome. The DAS more consistently distinguishes "repeat reassault" from each of the other four outcome categories, which explains why it has better overall prediction than the SARA. The fact that the DAS distinguishes "repeat reassault" from "one-time reassault" much better than the SARA makes sense, because it was developed

as a lethality instrument (to predict risk of lethal violence), whereas the SARA was designed to predict reoffending, not lethality.

For comparison to the risk instruments, the women's perceptions of risk were entered individually in multinomial logistic regression equations with no controls. Interestingly, women's perceptions of safety and how likely the man is to hit (assessed at intake) were also modest predictors of multiple outcomes by themselves, and were slightly better predictors than the simulated SARA. Feelings of safety at intake accurately classified 63% of "repeat reassault", which was almost as high a rate as the simulated DAS, although the rate of false positives was nearly 40%. Women's perceptions of how likely the man was to use violence in the next three months had a reasonably high sensitivity rate of 52% by themselves, with a relatively low rate of false positives (26%).

Women's perceptions had a higher rate of correct classification of repeat reassault than two of the three risk assessment instruments. However, there were more false positives with the women's feelings of safety as a predictor. The combination of the DAS and women's perceptions were the best models in this set of analyses (sensitivity = 63% and 64%; overall prediction = 35% and 37%). Nonetheless, they together still offered only modest prediction and were not as predictive as the initial multinomial equations with individual risk factors.

Hypothesis three was, therefore, supported, but the prediction was still not at clinically acceptable levels.

For comparison, we ran the same analyses with binary logistic regression, using "any reassault" over 15 months as the outcome (see Table 11). These analyses allow comparison to previous risk assessment research, which typically uses a dichotomous outcome. The patterns with regard to which risk assessment instruments and items provide the best prediction are virtually identical to the multinomial results. What is striking, however, is how much better the multinomial models do at predicting repeat reassault with these instruments and items than the binary models do at predicting any reassault. This confirms results discussed earlier, that

lumping one-time reassaulters with repeat reassaulters weakened prediction. Thus, our hypothesis that multiple outcomes would improve prediction using risk assessment instruments was supported. This finding strongly suggests that previous risk prediction has been weakened by not separating out the most serious reoffenders. Risk assessment may very well be improved by trying to identify and predict the most serious offenders.

BATTERER TYPES

As mentioned in the objectives, the third goal was to examine whether batterer types, as measured by the MCMI-III, are predictive of multiple outcomes. The four batterer types – “little or no pathology”, “narcissistic or antisocial,” “dependent/avoidant,” and “multiple disorders” – were entered as a set of dummy variables in the psychological characteristics block of the intake equations. The “no pathology” group was the omitted reference category. The batterer types variable was not a significant predictor of the multiple outcomes (or the dichotomous outcome) variable, when controlling for other significant predictors.

We also hypothesized that prediction might be improved by considering batterer type in conjunction with program characteristics. Batterer type, therefore, was interacted with program format (instructional versus discussion). Overall, the contribution of this interaction term to explained variance and risk classification (above and beyond the main effects equations) was negligible and not statistically significant. Likely, there were not enough differences in program delivery across the four sites to provide a powerful test of this hypothesis. The program formats, moreover, may not have been sufficiently distinct. The program approach across the programs, as indicated in the methods section, was fundamentally gender-based, cognitive-behavioral. The “format” refers more to how this approach was implemented, rather than to a different set of counseling assumptions. The formats also were not fully controlled or pure. There was some variation in the format among staff within a particular program. Future experimental research should be designed to specifically compare the interaction between

batterer type and program type, with interventions being tailored for specific types of batterers. Even so, we are not confident that prediction will improve given the instability in classifying batterers into specific types over time as recently acknowledged by Holtzworth-Munroe (2001), and the relatively even distribution of batterer types across outcome groups shown in our current study (Gondolf & White, 2001). Thus, hypothesis four, which predicted that batter type would be predictive of multiple outcomes and that an interaction term of batterer type by program type would improve prediction even more, was not supported.

CONCLUSION

Our prediction efforts demonstrated the importance of considering multiple outcomes in batterer research. Multiple outcomes modestly improved prediction and exposed different sets of predictors than dichotomous outcomes. Inclusion of conditional factors, however, did not substantially improve prediction. Future research with conditional models should predict shorter-term outcomes (i.e., over 3 months), rather than longer-term outcomes (i.e., over 12 months), which might show stronger support for conditional prediction models.

The use of simulated risk assessment instruments did not improve prediction over the models that used risk markers individually. One risk instrument (the DAS) did, however, offer prediction that was nearly as strong as the risk marker model. Another risk instrument (the SARA) offered modest prediction, while the third (the K-SID) offered very poor prediction of abuse outcomes. Importantly, women's perceptions of risk – their assessment of how likely the man is to use violence again and how safe they feel – were strong predictors of repeat reassault. In fact, they predicted repeat reassault better than two of the simulated risk instruments (SARA and K-SID). Risk assessment instruments, therefore, should be used with caution and should incorporate women's risk appraisals.

Our findings contradicted overgeneralizations about "high risk" batterers. They are not easily "typed" or predicted. Improvement of prediction and containment of the most dangerous

men will require further differentiation of types and/or abuse categories. In the next section, we explore a number of alternative batterer types and abuse outcomes to determine if they might help improve prediction.

PART VI: RESULTS OF FURTHER EFFORTS TO IMPROVE PREDICTION

INTRODUCTION

We investigated a number of other considerations that might help to improve prediction, especially the possibility of alternative batterer types and abuse outcomes. These alternatives might also help explain why our multinomial analyses did not more accurately predict abuse and reassault among batterer program participants. In this largely exploratory effort, we first examined the contribution of two sets of batterer types to prediction: racial groupings and types of psychopathy. Equations were computed separately for each racial group in our sample. We established psychopathy "types" using the MCMI-III and entered them as a risk marker in our multinomial equations as predictors. Second, an additional conditional variable is considered: perceptions of sanctions. Sanctions imply deterrence that might influence prediction of our abuse outcomes.

Third, we considered three alternative categorizations of abuse outcome. Using risk markers, we tested the prediction with non-physical abuse categories separately. We also identified patterns of violence based on women's narratives. These are compared to our current multiple abuse outcomes. We also reviewed our case material to identify atypical subjects and speculated on their impact on our prediction analyses. Lastly, we searched for variables associated with the women's perceptions of violence. Women's perceptions were one of the most consistent predictors in our multinomial analyses of abuse outcomes and appeared to warrant further examination. Understanding the women's "predictions" might help us improve prediction overall.

ALTERNATIVE BATTERER TYPES

Risk Predictions by Race

One alternative "type" of batterer we examined was based on racial groups. Separate prediction equations were developed for the three main racial groups represented in our sample: Whites, African Americans, and Latinos. We developed unordered multinomial logistic regression equations for Whites and African Americans using the previously discussed risk markers. However, because the sample size for Latinos was too small, a binary logistic regression equation was developed for this group. When running the models separately for African Americans, the ordered logistic regression models were not significantly weaker from the unordered multinomial models (based on the test for parallel lines and the likelihood ratios test described in the methods section). Nonetheless, we show results from the unordered logistic regression models for African Americans to assess the effects of specific risk markers (see Table 12) and for comparison to the unordered model for Whites (see Table 13). We present classification results for African Americans and Whites from the unordered multinomial equations (in Table 14) in order to compare prediction results with each other and with the intake models presented in the first results section. Tables 14 and 15 show results from the binary logistic regression equation for Latinos.

Overall, as examination of Tables 12 through 15 show, predictors for each racial group were more similar than dissimilar. The strongest prediction was obtained for Whites; the model was quite robust in predicting repeat reassault (79% sensitivity) and was stronger than the intake (70% sensitivity) and conditional models (57% sensitivity) computed for the racial groups combined. The prediction model was moderately strong for African Americans (65% sensitivity)

and comparable to the overall intake equation discussed earlier (70% sensitivity). The prediction model was weakest for Latinos (52%), and fewer risk markers were identified.

As Table 16 shows, running multinomial models with the risk assessment instruments separately by race was also instructive. The Danger Assessment Scale (DAS) worked best overall, and it was much more consistent across the three racial groups. The K-SID and SARA worked well with Whites, but not well with African Americans and Latinos. This finding may help explain why the K-SID and the SARA did not predict very well in the overall equations discussed earlier. Women's perceptions of safety by itself was a solid predictor of repeat reassault for Whites, but not for African Americans and Latinos. Women's perceived likelihood of violence predicted repeat reassault best among Latinos, but was also a solid predictor among Whites and African Americans.

There were similar patterns using binary logistic regression to predict any reassault (versus no reassault) using the risk assessment instruments and women's perceptions across the 3 racial groups. Again, DAS worked best overall, with similar results in all 3 racial groups. The K-SID fared poorly in all 3 racial groups, and the SARA predicted reasonably well among Whites but not among African Americans and Latinos. Perceptions of safety was a solid predictor of any reassault among Whites and African Americans, but a weak predictor among Latinos. Perceptions of how likely the man was to use violence was a strong predictor of any reassault among African Americans, solid among Whites, and a very poor predictor among Latinos.

Conclusions

Prediction of abuse outcomes does appear to be affected by racial background and may be improved by developing separate prediction for different racial groups. Much risk assessment, including development of risk assessment instruments, has been done with predominantly White samples. Latino samples of batterers are quite rare. This fact is clearly

reflected in the results, with the best prediction occurring among Whites and the weakest prediction among Latinos. We recommend that, to the extent that criminal justice personnel and practitioners use risk assessment instruments, they should use them with extreme caution when working with minorities, especially Latinos. The results also suggest that larger samples of minority men will be needed to develop more refined risk assessment instruments. It is, however, premature to recommend that separate risk assessment instruments have to be developed for separate racial groups.

In sum, our hypothesis that separate prediction for different racial groups would improve prediction with abuse outcomes was partially supported. There were mostly similar risk markers across the three racial groups, but some different ones appeared as well. The sensitivity of the prediction for Whites alone was improved over the prediction for the combined sample, and approximately the same for the other two racial groupings. Prediction using risk instruments was noticeably varied among racial groups.

Batterers with Psychopathic Tendencies

We explored for additional batterer types based on psychopathic tendencies. Psychopathy has received a good deal of attention in the efforts to understand violent men and may be the key to identifying the most unrelentingly and severely violent men. It is considered to be an improvement of the broader and imprecise conception of an antisocial personality (Cunningham & Reidy, 1998). The conventional conception of psychopathy is that the cold-hearted and unfeeling tendencies are related to brain dysfunction. These men tend to be particularly dangerous and unresponsive to treatment (Cleckley, 1988; Hare, 1996). This "primary" psychopathy conception views psychopathy as a discrete personality problem.

You either have it or you don't (Cleckley, 1988; Hare, 1996). Some experts propose, however, a "secondary psychopathy," in which anxiety appears along with antisocial and sadistic tendencies (Blackburn, 1996, 1998; Richards, 1993, 1998). This latter grouping of

highly distressed individuals is more likely to be responsive to intervention and treatment (Melow, 1988). In this view, it is also possible to experience a range of psychopathic tendencies – from a personality “disorder” to a less severe form of a personality style (see Gondolf & White, 2001, for a full discussion of this conception and its implications).

Profile Grouping Procedure

We investigated the contribution of an additional batterer type by interpreting further the men’s MCMI-III profiles. We followed the profile interpretations recommended in the instrument manuals with one revision. We identified psychopathic tendencies according to profile configurations recommended by experts on psychopathy and the MCMI (Blackburn, 1998; Millon & Davis, 1998)(see Gondolf & White, 2001, for a detailed description of the classification procedures). Any evidence of psychopathic tendencies was given priority over other possible interpretations in order to ensure the maximum inclusion of such tendencies. The broader and more liberal conception of psychopathy is likely to identify more men than narrower conceptions previously used in the field. In other words, we separated any profile with psychopathic tendencies from the standard profile groupings used in our previous analyses of batterer types. (Psychopathy is currently not considered to be a distinct psychiatric disorder, and it therefore does not emerge in our previous batterer types based on the DSM-IV). Overall, as Table 17 shows, we found that profiles for 9% of the men suggested conventional or “primary” psychopathic disorder and 11% suggested “secondary psychopathology. The broadest possible conception of psychopathy (which we called “psychopathic tendencies”), including secondary psychopathy and both psychopathic “disorder” and “style,” applied to 42% of the men overall.

The new batterer types – “primary psychopathy,” “secondary psychopathy,” and “psychopathic disorder or styles” – were constructed for all the cases having complete information for the 15-month follow-up (n=580). The psychopathy types were first cross-tabulated by three categories of reassault (i.e., no reassault, one-time, and repeat) in order to

test the bivariate association of psychopathy to reassault. Were those who repeatedly reassaulted more likely to be psychopathic? Second, we tested the contribution of the psychopathic types to prediction of the multiple outcomes (i.e., no abuse, verbal abuse/control, threats, one-time reassault, and repeat reassault) using the multinomial procedures previously applied to risk markers and conditional variables. We entered primary psychopathic disorder, secondary psychopathic disorder, and psychopathic disorder or style separately as dummy variables into the intake and conditional prediction equations.

Major Findings

Based on previous research, we expected batterer program participants who repeatedly reassaulted to be characterized by psychopathic tendencies and to be significantly more likely to have these tendencies than participants who did not reassault their partner or did so only once. Approximately half of the men who “repeatedly reassaulted” showed some evidence of psychopathic tendencies in the broadest sense of psychopathy (see Table 17). However, a relatively small portion (1-in-10; 11%) of men in the “repeat reassault” category exhibited primary psychopathic disorder—the classic “cold-hearted” psychopathy of greatest concern.

The men who repeatedly reassaulted reflect a diversity of levels and types of personality and psychopathology. Nearly two-thirds (60%) had “subclinical” or low levels of personality dysfunction. The findings do not substantially differ across the three different reassault outcomes (i.e., none, once, repeated). The men in the “repeat reassault” category are no more likely to appear with a psychopathic disorder than those men in the other reassault categories. It appears difficult, therefore, to typify or generalize about the personalities of the demonstrated “high-risk” or “dangerous” cases.

None of the psychopathy types were predictive of multiple outcomes, when other significant predictors were controlled. Neither “primary psychopathy” nor “secondary psychopathy” were significant bivariate predictors of multiple outcomes. Although having any

psychopathic tendencies was a statistically significant bivariate predictor of multiple outcomes, it was not a strong predictor and did a poor job of correctly classifying cases into categories, including repeat reassault.

We also assessed for an interaction effect between race and psychopathy. We did so by ascertaining whether the psychopathy measures were significant risk markers in the equations that were computed separately for the three racial groups. None of the measures of psychopathology were predictive in any of the equations by race, once other significant predictors were controlled.

Our hypotheses regarding psychopathology were not supported. It appears that batterers with psychopathic tendencies are not more likely to repeatedly reassault, either overall or within racial groups.

Additional Conditional Variable: Perceptions of Sanctions

We explored an additional conditional variable that has been suggested by deterrence theory. We tested the influence of men's perceptions of sanctions on the prediction of both dropout and reassault. We first explored how men derive their perceptions of sanctions in the form of an "experiential effect." Will batterers perceive sanctions as likely based on the imposition of previous sanctions? Further, will batterers with prior justice system contact be more likely to perceive sanctions as likely? We then explored for a specific deterrent effect, which refers to the direct preventive effects of legal sanctions imposed on individuals who actually engage in illegal behavior. Since almost all of these men have been arrested, specific deterrence would predict that men who perceive sanctions to be likely should be deterred from further violence (Akers, 1997; National Institute of Justice, 1999).

To explore for an "experiential effect," we used binary logistic regression to estimate program and batterer characteristics that predict the perceptions of sanctions. We then explored for a "specific deterrence effect" by using binary logistic regression to determine

whether the type of batterer who perceived legal sanctions as likely would also predict program dropout or reassault.

Methods

Our analyses required two outcome measures: program dropout and reassault during the follow-up period. As discussed in the Measures section, "program dropout" was defined as attending less than 12 sessions. A second dropout variable, "dropout before completion", was coded '1' if the batterer did not attend the total number of sessions required by each program. This number varied from 12 at Pittsburgh and Dallas to 22 in Houston and 36 in Denver. The first dropout variable is used in our analysis for predicting perceptions and reassault for batterers at all four sites. The second dropout variable is used as an outcome variable (to determine if the subtype who perceived sanctions as likely were less likely to dropout), although men from Pittsburgh and Dallas had to be excluded for this analysis (because the program had ended).

The second outcome variable we used was "any reassault", which was coded 1 if any of the women (initial or new partners) or the man reported any assault as occurring between 3 months and 15 months. Reassaults measured at the 3-month follow-up were omitted because men's perceptions of legal sanctions were also measured at the 3-month follow-up and could have been influenced by a reassault between the intake interview and the 3-month follow-up and its attendant consequences (i.e., experiential effect).

The perceptions of sanctions were assessed at the 3-month follow-up (3-months after program intake) to capture the possible effect that program contact may have added to the perceptions beyond the batterer's arrest and court experience. In the initial program sessions at each program, staff instructed the participants about the fact that domestic violence is a crime and about the designated legal sanctions for program dropout and reassault.

Procedures

To identify predictors of perceptions of certainty of sanctions, two logistic regression equations were developed – one for dropout sanctions and one for reassault sanctions. Because of a bimodal, skewed distribution, the two “likelihood of jail” variables (one for dropout and the other for reassault) were converted into dichotomous variables (i.e., men perceiving jail as unlikely or uncertain versus men perceiving jail as likely). For both perceptual variables, the men’s social background variables were entered first, followed by the relationship variables, the prior behavior variables (e.g., alcohol/drug use, system contact, and other abuse variables), personality variables, program context (i.e., program site and referral source), and post-intake variables (i.e., program dropout and committing physical reassault between intake and the 3-month follow-up). Race and age were entered and left in both equations as control variables.

To determine whether perceptions of the certainty and severity of sanctions were predictive of reassault between 3 months to 15 months after intake, a similar procedure was used. Identifying appropriate control variables was achieved by entering race, age, and the same blocks of variables described above in a hierarchical fashion. In addition, no contact and reassault between intake and the 3-month follow-up were included as additional control variables. Once other relevant determinants of reassault were identified through this hierarchical stepwise procedure, the four risk perception variables—likelihood of jail for dropout, likelihood of jail for reassault, most severe sanction perceived for dropout, and most severe sanction perceived for reassault—were entered individually, one at a time.

The equation for dropout between the 3-month follow-up and the 15-month follow-up was developed in exactly the same fashion as the reassault equation. However, two sites, Pittsburgh and Dallas, were omitted from the dropout equation because these programs were completed at the time of the first follow-up. Thus, temporal order of program dropout and sanctions perceptions at the 3-month follow-up cannot be established for these two sites. The men at these sites would already have completed the program when the perceptions were

assessed. Therefore, the batterers from Pittsburgh and Dallas were not at risk for dropout after the first follow-up.

Results

Logistic regression results for the likelihood of jailing for dropout and reassault are reported in Table 18. There was clear evidence of an "experiential effect." Nearly half of the men perceived sanctions as likely if they dropped out of a program or reassaulted. Further, variables that measured prior justice system contact were significant predictors of the men's perceptions. Although the explanatory power of these multivariate equations is adequate based on Chi Square statistics ($39.1_{df\ 14}$ and $75.8_{df\ 15}$; $p < .001$), the predictive power of the equations (i.e., the ability to correctly identify perceptions) is weak (overall correct classification of 63.3% for reassault and 70.3% for dropout). Table 19 shows results for the adjusted effects of both certainty and severity of sanctions on dropout and reassaults occurring between 3 months and 15 months, while controlling for other significant predictors. There was no evidence for a "specific deterrence" effect. Neither certainty nor severity variables had a significant impact on the likelihood that men would dropout or reassault.

Discussion

The results substantiated an "experiential effect." A substantial proportion of the batterers did perceive jailing as a likely sanction for dropping out of a program or reassault. There was no evidence, however, of a "specific deterrence" effect. Batterers who perceived sanctions as certain and severe were not less likely to dropout or reassault (see Heckert & Gondolf, 2000a for possible explanations and policy implications). These results suggest that swift and certain punishment for program dropout or reassault may not in itself be enough to reduce spouse assault by men. Batterer counseling must, therefore, account for at least some of the reduction in abuse observed in our previous studies of program outcome (Gondolf, 1997b). The

relevance to our current prediction research is that “perceptions of sanctions” do not appear to improve prediction. Our hypothesis that “perceptions of sanctions” as a conditional variable would improve prediction was not supported.

ALTERNATIVE OUTCOME CATEGORIES

Risk Markers for Non-Physical Abuse

We attempted to identify risk markers and improve prediction for alternative categories of non-physical abuse. Our multinomial results suggest that non-physical abuse is a correlated but distinct phenomenon from physical abuse. In addition, there are different risk markers that distinguish threats from repeat reassault and other forms of emotional abuse from repeat reassault (O’Leary, 1999). These previous findings raise the question of whether specific forms of non-physical abuse might be easier to predict.

The Measures section describes the three types of non-physical abuse assessed at intake: verbal abuse, controlling behavior, and threats. To examine non-physical abuse as an outcome, we created a binary indicator for each type of non-physical abuse – that is, if the woman reported *any occurrence* of each of these types of abuse at *any* of the five follow-ups. We also computed a binary “severe non-physical abuse” variable indicating that the batterer had used all three types of non-physical abuse at least once during the 15-month follow-up. To adjust for possible under-reporting by the women, the women’s reports of no abuse were changed to abuse if a batterer report was available and indicated abuse. We used the same predictor variables that were used for our multinomial analysis of multiple abuse outcomes including reassault. We computed four binary logistic regression equations to identify risk markers associated with the four different categories of non-physical abuse during the 15-month follow-up (additional details are available in Gondolf, Heckert, & Kimmel, in press).

Results

The results for the logistic regression equations used to identify risk markers of recurring non-physical abuse are shown in Table 20. The only consistent risk markers of non-physical abuse were "the use of non-physical abuse prior to program intake," "use of severe physical abuse prior to intake," and "program attendance". Men who used non-physical abuse and severe abuse prior to intake were more likely to use non-physical abuse in the follow-up. Program dropouts (by 3-months) were also more likely than program attenders to abuse their partner non-physically during the follow-up. Batterer demographics, family-of-origin abuse, heavy drinking, and prior antisocial behavior were not substantially or consistently associated with non-physical abuse. Also, the program sites were not significantly associated with any of the abuse categories. The equations for different categories of non-physical abuse suggest, moreover, different patterns of risk markers for different categories of non-physical abuse. For instance, risk markers for controlling behavior included several demographic characteristics (e.g., ethnicity, marital status, and occupation) but did not include abuse in the family-of-origin. Verbal abuse, on the other hand, included abuse in the family-of-origin, but no demographics.

As shown in Table 20, overall prediction of the non-physical abuse categories was modest (see overall classification rates, which range from 65% to 81%), and quite comparable to the overall classification of "any reassault" (75%) in the dichotomous prediction equations presented earlier. In addition, with the exception of verbal abuse (97% sensitivity), sensitivity rates were low (control = 67%; threats = 65%; severe = 32%), although all but one was higher than the sensitivity rate for "any reassault" (44%). Thus, prediction of non-physical abuse is only marginally improved over the multiple outcomes model by treating them as distinct outcomes.

Discussion

In sum, our hypotheses that additional categories for non-physical abuse would reveal additional risk markers and improve prediction were not supported. Past abuse, once again, was a substantial predictor for continued or recurrent abuse, and program participation appears, at least to some degree, to reduce this recurrence. Non-physical abuse in our clinical sample of batterer program participants had fewer consistent risk markers than in general population samples (e.g., Follingstad et al., 1990; Margolin et al., 1998; Stets, 1991; Sugarman et al., 1996). A similar finding has occurred in attempts to predict physical abuse (Jones & Gondolf, 2001), which suggests that more detailed calibrations may be necessary to develop risk markers in clinical samples.

Violent Incidents

We explored for alternative outcome categories by qualitatively coding victims' descriptions of violence. Our research has been unable to identify differences between the men who repeatedly reassault and the other men based on batterer type, psychopathic tendencies, or other psychological characteristics. One possibility is that these categories of reassault need to be delineated further. There may be additional or more precise categorizations to consider. Another possibility is that a different categorization of violence may be easier to predict. Our multiple outcome categories are based on cumulative behaviors. Some violence researchers have argued that violence is a process rather than an isolated behavioral event, and that consideration of this process might produce different outcomes (Monahan, 1996; Mulvey & Lidz, 1993; Steadman, 1982). Our research question is whether the one-time reassault and repeat reassault categories are distinguishable by the process of the physical abuse? Or are other categorizations of violence evident?

Methods

To answer these questions, we coded the women's descriptions of the violent incidents, categorized those codes, and cross-tabulated those categories with our multiple abuse categories that include one-time versus repeat reassault. The men's partners were asked to describe in detail the violent incidents they reported. The resulting narrative accounts offered extensive detail on the process and circumstances of the violence. As stated in the Measures section, we coded the narratives using a sequential, situational process model of violence, recommended by other violence researchers (Monahan, 1996; Mulvey & Lidz, 1993; Steadman, 1982). Violence, in this view, begins with the circumstances and issues of the incident, the interaction and dynamics of the incident, and ends with the man and woman's response to the incident.

The coders first identified the issues, circumstances, precipitants, alcohol use, man's emotional state, couple interaction, pattern of tactics, and responses of each incident of violence. There were a total of 536 incident descriptions among the 299 women who responded for the full 15-month follow-up. These included reports of the incidents preceding program intake (i.e., one for each of the 299 men) and 237 incidents described during the follow-up. We also summarized the course of violence over the incidents that an individual woman may have experienced. There were four major components in this case-level coding: "women's issues," "batterer problems," "incident pattern," and "violence type." (See Gondolf & Beeman, under review for further details.)

To answer the research questions, we cross-tabulated each component derived from our coding by the previous reassault category and looked for tendencies across the responses. We noted items that varied 10 percentage points or more across the reassault categories. Significant levels are not technically appropriate in this instance given the qualitative basis of the data and the absence of true random sampling.

Results

We were not able to find a distinguishing pattern or type of violence for the repeat reassault category. In other words, our coding of the process of violence did not correspond with our behavioral-based categories of reassault. As Table 21 reveals, there were few differences in the issues, precipitants, circumstances, or emotions of incidents across the categories of reassault. The men who repeatedly reassaulted were, moreover, not more likely to be the source of the more severe patterns and dynamics of physically abusive incidents.

Men who repeatedly reassaulted were slightly more likely to use a chain of tactics, or multiple tactics, in their violent incidents (9%). That is, their violence was slightly more likely to be excessive and unrelenting than men with one reassault (4%). As Table 22 shows, overall there was a small subcategory of men (4%) who committed unrelenting and excessive violence across all the reassault categories. The men who repeatedly reassaulted were also more likely to be possessive and controlling, as shown in Table 23.

The most substantial difference between batterers in the one-time and repeat reassault categories was in the women's interaction during, and response after, the violence. As can be seen in Table 23, the partners of the men in the repeat reassault category were less likely to resist the violence during an attack and less likely to seek help in response to the violence. These women may be less assertive because they are more fearful of retaliation from consistently abusive and controlling men, as Jacobsen & Gottman (1998) suggest in their laboratory study identifying "cobra" and "pit bull" types of batterers. These women may also be discouraged by the failures of their previous help-seeking efforts and feel their efforts to stop the violence are not worthwhile. As a result, the repeat reassault continues unchecked.

The women's lack of assertiveness may, furthermore, reflect the dynamics of the relationship in general. The broader context of non-physical abuse and daily control may so entrap or debilitate some women that they are unable to assert themselves. Examination of the relationship in general, rather than focusing only on incidents, might further distinguish the

reassault categories. In any case, women's assertiveness appears to warrant more consideration in the effort to distinguish, predict, and contain repeat reassault.

Discussion

Overall, the hypothesis that there may be alternative abuse outcomes received some support. The mode or pattern of violence, however, is not associated with our cumulative categories of reassault (i.e., none, one-time, and repeat reassault). Additional development of these or other conceptions of violence may prove fruitful for future prediction studies. Of special note is a category of excessive, unrelenting violence, which is a small but extremely dangerous outcome. While it was slightly more likely to coincide with our repeat reassault category, it was present in the other categories as well.

Results for Atypical Cases

We conducted a qualitative analysis to identify "atypical" categories of batterer "types" and abuse outcomes that might decrease our ability to accurately predict reassault. There is diversity among batterers and their partners that may be difficult to capture in categorical measures of characteristics and statistical equations. Recent qualitative and clinical studies of battered women and of batterers show a range of violence and personal circumstances. Perhaps a search for patterns among these diversities might capture abuse outcomes that conventional measures of reassault categories, demographics, personality, and past behavior cannot.

Methods

As discussed earlier, interviewers nominated "atypical" cases at weekly meetings, which identified any man, woman, or couple who were distinct, unique, or exceptional in terms of their characteristics or circumstances. Two research assistants, who were not involved in the

interviewing, read through the interview records for each atypical case and wrote a case summary. The case summaries were then grouped according to the distinguishing features of the case.

Approximately 120 cases were classified as "atypical" out of those who were interviewed for at least the full 15-month follow-up (n=561). This amounts to over 20% of the available cases. Most of these cases were distinguished by the nature of the violence during the follow-up, such as severe child abuse and molestation (n=11), excessive violence including marital rape (n=20), violence against multiple partners (n=7), problems with getting needed services or help (n=10), and severe violence from the woman (n=4). At the other extreme, several cases (n=11) were identified as resounding "successes" (see Gondolf, 2001, for additional details).

Results

Some cases were noteworthy because of their exceptional characteristics assessed at the time of program intake (n=34). The most interesting are those with atypical careers. One woman, for instance, was a lawyer whose caseload was primarily battered women, and another was a "star" in pornographic films. One man was a professional basketball player in Europe, and another was an international drug smuggler with private planes and yachts. All of these cases encountered difficulties accessing conventional battered women services or batterer counseling. The women risked a loss of anonymity in the women's services, or they did not "fit in" to the services. The men were, on the other hand, able to avoid the batterer programs by going to their international residences or using their wealth to "buy" an alternative to counseling.

A few men and women had severe health or psychological problems. One man was struggling with a terminal disease, a few women were suicidal, and several women and men reported being in psychiatric treatment for serious depression. Chronic drug and/or alcohol addiction marked other cases (n=20). Another woman described her uncontrollable outbursts of rage and violence. Her husband was defending himself against her violence toward him. For

these men and women, the health or psychological problems were the primary concern and warranted specialized treatment.

Some relationships were also very complicated and even chaotic. One man was involved with several partners at one time, and abusive toward them all. Another case included a husband assaulting and raping his wife, the grandfather molesting their teenage daughters, and one of the sons being violent towards them all. Some families included foster children, stepchildren, older relatives, and even sexual "triangles." Many of these cases require intervention with more than just the arrested "batterer," or further investigation into the actual circumstances of the arrest.

Discussion

Our hypothesis that the atypical cases would identify diversities and categorizations that could reduce prediction was supported. The atypical cases may be another reason our prediction efforts are still relatively weak. It is difficult to account for the range of diversity among clinical samples of batterers and their victims, and there may be much more of it than our current generalizations suggest. Some of this diversity may be addressed in refinements of batterer types or abuse outcome categories – that is, by making our analytical models even more complex. Prediction might be improved, on the other hand, by eliminating some atypical cases as outliers and developing prediction with the "normative" cases. The latter may be a concession to the fact that an amount of unavoidable unpredictability is inevitable. The question is how much of such "unpredictability" can we accept in clinical situations and how do we respond to it.

Our qualitative review of case characteristics identified 20% of subjects who could be considered atypical in terms of careers, mental health, relationships, or violence. Some men warrant special attention within a group, and others may benefit from referral to additional treatment. Some women, on the other hand, would seem to warrant much more counseling and

support of their own given the severity and duration of past abuse, and few obtaining domestic violence services in the past.

Determinants of Women's Perceptions of Risk

Another area to consider in the attempt to improve prediction is the women's prediction of risk. Researchers have long argued that clinical judgment is less accurate in predicting violence than risk markers of actuarial information (Grove & Meehl, 1996). However, we found that women's perceptions (i.e., "judgments") of how likely it is the man will use violence and their perceptions of safety were consistently predictive of the use of threats and reassault, including repeat reassault, above and beyond other risk markers. In addition, women's perceptions *alone* (with no covariates) were more predictive of repeat reassault than simulated versions of two of the three popular risk assessment instruments. This finding is impressive considering that these perceptions were measured at intake and significantly predicted cumulative multiple outcomes over fifteen months.

Presumably, identification of determinants of women's perceptions could help researchers and practitioners improve their predictions of reassault. Yet, there has been little research regarding how women arrive at their risk assessments (Langford, 1996). The overall objective of the analysis in this section is to identify the basis of women's perceptions of risk. What variables are associated with women's perceptions of how likely the man is to be violent and how safe the woman feels? These variables should offer some indication of the basis or criteria of women's predictions. This information could help us refine risk markers and instruments or the prediction analysis we have attempted to develop.

Results

Tables 24 shows the univariate distributions of the women's perceptions measures. Clearly the majority of the women feel safe (89%), *at intake*, and believe that it is unlikely the

batterer will use violence over the next three months (68%). Table 25 shows the bivariate relationship between women's perceptions of the likelihood of violence and their perceptions of safety. There is a moderate to strong association between the two variables (Somer's $D = .49$). In fact, perceived likelihood of violence is the single best, bivariate, predictor of perceptions of safety. Although they covary, we argue they are measuring different constructs, as they do function independently in the multivariate risk assessment equations.

Multivariate Analysis: Because the two dependent variables, women's perceptions of risk, are ordinal, we estimated two multinomial logistic regression equations for each variable, using cross-sectional data collected at baseline. First, we developed an unordered multinomial (or nominal) logistic regression model, and then we compared the results to an ordered logit model (proportional odds model). For both perceptual variables, the men's social background variables were entered first into the model, followed by the relationship variables, the prior behavior variables, psychological characteristics, and women's drinking and use of physical violence. We used a stepwise approach by testing the effect of each predictor in each block one at a time, while controlling for race, age, program site, referral source, and any other predictor variable that had achieved a two-tailed significance of $p \leq .10$ (see Hosmer and Lemeshow, 1989).

Women's perceptions of how likely the man is to use violence again was entered as the last determinant in the women's perceptions of safety model. We view perceptions of violence likelihood as a perceptual variable that mediates the relationship between the other determinants and women's perceptions of safety. Credit for any mutually shared explained variance in women's perceptions of safety was, therefore, given to determinants other than perceived likelihood the man would use violence again in the next three months.

Comparison of the ordered logistic regression models to the unordered multinomial models showed few differences. We were also unable to reject the null hypothesis of parallel lines, and the likelihood ratio tests were not significant. In other words, there was not sufficient

evidence to require use of the unordered results. Accordingly, results of the ordered logistic regressions are presented in Tables 26 and 27 for ease of interpretation. We also present the standardized regression coefficients from the same models estimated with ordinary least squares regression. The results of the OLS equations were similar to the results from the ordered logistic regressions, and the standardized regression coefficients provide some indication of the relative importance of the independent variables in determining women's perceptions of risk.

As revealed in Table 26, the final model for women's perceptions of how likely the man is to be violent during the next three months was modest ($R^2 = .25$). Eight determinants of violence likelihood were significant at a two-tailed alpha level of .10 in the OLS multiple regression equation, although all were relatively modest predictors according to the size of the standardized OLS regression coefficients. Significant determinants of women's perceptions that the man was likely to use violence were the man had been frequently drunk or high in the last year, voluntary status in the program, unemployment, controlling behavior, use of threats, being divorced or separated, the man had ever used severe violence, and the woman was frequently drunk or high in the last year.

As Table 27 demonstrates, the overall model for women's perceptions of safety was stronger because of the inclusion of women's perceptions of the likelihood of violence in the model ($R^2 = .40$). By far, the strongest determinant of women's perceptions of safety was their perception that the man was likely use violence during the next three months. Other significant determinants for feeling unsafe were program site, the man was frequently drunk or high in the last year, the man was married, the man had used controlling behavior in the last three months, the man and woman were not living together, and the man had ever used severe violence.

In sum, the best predictors of women's perceptions of risk were previous use of psychological and physical abuse, relationship characteristics, and drinking and drug use. Women appear to base their risk predictions on many of the same risk markers identified in the

prediction research. Clearly, however, there are other factors that influence their perceptions that are not typically measured in the research on risk markers and risk instruments, and that are missing from our database.

Qualitative Analysis of Women's Perceptions: Although the analyses in this report focus on the first 15 months of follow-up, a sample of the women and men in the multi-site evaluation were interviewed every 3 months for 48 months. At 33 and 36 months after intake, the women were asked an open-ended question about their reasons for their perceptions of safety and predictions of reassault, which a graduate student thematically coded. In summary, the women's perceptions and predictions appeared to be based primarily on their immediate observations of the man's behavior (e.g., "he was less abusive" or "appeared to have changed") and consideration of current circumstances (e.g., no longer contact with the partner or partner was in jail). Specifically, about one-third of the women cited lack of contact with the batterer, about one-third noted change in their partner or in the relationship, and about one-fifth believed that the initial abuse was a one-time event. The rest of the women saw no change and did not feel safe.

We cross-tabulated these qualitative reasons with sociodemographic variables and past behavior (results not shown). Two types of significant factors were relationship status (unmarried, having no children, and living apart felt safer) and previous abuse. In addition, if the man admitted he had a problem at intake (according to the woman), the woman was more likely to identify him as changed.

The qualitative analysis clearly reinforces the quantitative findings. Previous abuse and relationship characteristics are the primary factors that women consider when forming their appraisals of risk.

Discussion

Overall, the hypothesis that the best predictors of women's perceptions would be relationship status, alcohol use, and previous abuse, was supported, although the prediction was modest. Women's perceptions appear to be based on rational and practical considerations. According to the multivariate analysis, they rely primarily on relationship status and past abuse as rational guides to their risk appraisals. These factors are also risk markers for reassault, which suggests that many of the women are cognizant of factors which place them at higher risk for assault.

The women appear to rely on conventional risk markers, yet they attain equivalent or better prediction than some risk instruments. The most basic explanation is that the women simply have more information about abuse available to them. They observe first hand a "constellation of abuse," rather than the isolated components we measure in our analyses (Dobash & Dobash, 2000; Dobash et al., 2000). A second explanation is that the women "predict" in an on-going process of identification and response to a variety of cues, as one descriptive interview study suggests (Langford, 1996). That is, they are involved in an iterative process that is far more complex than our static, or even conditional, risk marker models suggest. Another explanation is that they are able to develop idiosyncratic models of prediction that fit their particular batterer. Each batterer may have his own unique pattern or trajectory of behavior.

Future prediction research is likely to benefit from further examination of how women "predict" violence. The process of women's prediction needs first to be described. Second, lessons from the process need to be transferred to clinical judgments and risk instruments.

CONCLUSION

Our exploratory analyses to identify additional or alternative conceptions for prediction models surfaced a few possibilities to consider in future research. Prediction is modestly

improved for Whites alone, as opposed to a combined sample of races, which suggests a need to develop separate risk instruments for different racial groups. Prediction for non-physical abuse is marginally improved over the multiple outcomes that include reassault. We also found some evidence that patterns of violence, rather than our cumulative abuse outcomes, may warrant investigation, and that nearly 20% of the subjects could be considered as atypical in terms of their careers, mental health, or violence. Interestingly, types of psychopathy and perceptions of sanctions did not contribute to prediction of multiple outcomes, despite speculations in the field that they would. Lastly, we tested for predictors of women's perceptions of future abuse and safety, since these perceptions were a consistent predictor in our multinomial analyses. The women appear to draw on the prevailing risk markers, but apparently use a more complex process of assessment to improve their predictions. These exploratory analyses point to accounting for the diversity of batterer backgrounds and further delineation of violent events in prediction, and for the process of prediction suggested in women's perceptions. The question remains whether models can be developed to adequately account for these complexities and whether those models can be effectively implemented in clinical settings.

PART VII: CONCLUSION

INTRODUCTION

In this part of the report, we first summarize our findings pertaining to our primary and secondary objectives. Next, we discuss a number of limitations and qualifications of our research. Finally, we discuss various implications for future research and clinical practice.

SUMMARY OF FINDINGS

The primary objective of our research was to improve prediction using a more complex analysis that included multiple outcomes and conditional factors. Specifically, we attempted to identify risk markers, risk instruments, and batterer types that help predict emotional abuse, reassault, and especially repeat reassault. The secondary objective was to identify additional ways to improve prediction by exploring alternative subtypes of batterers and alternative categories of abuse. Predicting reassault and "dangerous" cases has become an important concern in the field of domestic violence. Accurate risk assessment is essential to better protect victims and to assign batterers to the appropriate interventions.

Our attempt to improve prediction of further abuse by batterer program participants produced some instructive findings. As hypothesized, using multiple outcomes does improve prediction. When using intake risk markers, the multiple outcome model successfully predicted 70% of the "repeat reassaulters," whereas the dichotomous outcome models (reassault versus no reassault) only predicted 44% of "any reassaulters." This finding clearly suggests that researchers and program evaluations need to distinguish between one-time reassaulters and repeat reassaulters, especially when doing risk assessment. The overall multiple outcomes model, however, was modest; only 48% of the batterers were correctly classified across the five outcomes. Overall prediction would probably improve by collapsing the "no abuse" and "controlling/verbal abuse only" categories. Significant risk markers were fairly consistent with

previous research. Importantly, women's perceptions of risk emerged as an important predictor, whereas men's psychological characteristics did not, net of other significant predictors.

The hypothesis that prediction would be improved by including conditional or time-varying variables into the model was not supported. The overall conditional model ($n = 399$) was fairly robust, but the ability to correctly classify (or predict) repeat reassaulters did not improve, relative to the static model ($n = 499$). Nevertheless, some conditional factors did emerge as important risk markers and slightly improved prediction (within the conditional models: $n = 399$), suggesting that future research should incorporate dynamic factors, especially explanatory or causal research. Because our conditional models predicted long-term multiple outcomes (i.e., cumulative outcomes between 3 to 15 months), conditional models that predict short-term outcomes (e.g., non-cumulative outcomes over 3 months) should be developed in future prediction research to assess whether they would improve prediction beyond intake models. Such research, using more dynamic models, is essential to better evaluate the utility of "on-going case management."

We also attempted to improve prediction by simulating three popular risk assessment instruments. One of the simulated risk assessment instruments (DAS) was a reasonably strong predictor of repeat reassault, one (SARA) was a modest predictor, and one (K-SID) was a very poor predictor. Only the Danger Assessment Scale (DAS) was more predictive than were the women's perceptions of risk by themselves. The DAS predicted "repeat reassault" almost as well as the intake risk markers entered separately, but it had twice the rate of false positives. Thus, the strongest prediction occurs by entering risk markers as individual items (and including women's perceptions of risk), rather than combining them into a composite index, especially for Whites. This does suggest that composite indices could possibly be improved by weighting items based on their relative contributions to the prediction models. Importantly, the risk instruments did much better in the multiple outcomes models predicting "repeat reassault" than in the dichotomous outcome models predicting "any reassault."

Consideration of batterer types did not significantly improve prediction, either as a main effect or when interacted with program format. Antisocial batterers were not more likely to reassault or repeatedly reassault than other types of batterers. Further, whether or not the different types of batterers attended instructional or discussion oriented programs did not increase or decrease their risks of repeat reassault.

We also attempted to improve prediction by developing separate prediction equations by race. Using intake risk markers, we achieved very good prediction of repeat reassault (79%) among Whites, modest prediction among African Americans, and poor prediction among Latinos. With regard to the risk instruments, the DAS worked best overall and achieved consistent prediction across the three racial groups. The SARA and K-SID instruments worked reasonably well for Whites, but they worked poorly for the other two racial groups. Women's perceptions of risk were also modestly predictive within racial groups. Clearly, to improve risk assessment, additional research with larger samples of minority batterers is necessary. It will be especially important to develop and test risk instruments among African Americans and Latinos.

Further attempts to identify batterer types by developing measures of psychopathology did not improve prediction. We identified psychopathic tendencies according to profile configurations recommended by experts on psychopathy and the MCMI. Batterers who repeatedly reassaulted, however, were not significantly more likely to be primary psychopaths, secondary psychopaths, or to have psychopathic tendencies.

For the most part, the programs were successful at instructing the men that sanctions would be likely if they dropped out or reassaulted. Contrary to expectations, however, identifying the subtype of batterer who "learns" that sanctions are likely did not improve prediction of reassault outcomes. Swift and certain punishment for program dropout or reassault may not be enough to reduce spousal assault by batterers. It may be important,

therefore, to enhance and improve the treatment component of batterers programs to improve their effectiveness and to take other measures to ensure victim protection.

We explored a number of categories of non-physical abuse to see if prediction could be improved by modeling them separately from physical reassault. The best predictors of non-physical abuse were previous non-physical abuse and previous physical abuse. In addition, batterer program treatment did appear to reduce non-physical abuse somewhat. There were no other consistent risk markers across the different types of non-physical abuse, and overall prediction of non-physical abuse was relatively modest. Correct classification of batterers who used various forms of non-physical abuse was also modest. Therefore, prediction will probably not be improved by developing separate models for different types of psychological and physical abuse; use of multinomial models with multiple outcomes is sufficient.

Qualitative analysis of violence narratives did not substantiate our hypothesis that there would be an alternative categorization of outcomes that might substantially improve prediction. There did appear to be a subset of men in the "repeat reassault" category (9%), however, which engaged in unrelenting, escalating, and coercive battering. Future prediction research may be more successful if this category of battering can be successfully identified and predicted. Identification of "atypical" cases reaffirmed our other findings; there is a remarkable diversity across cases, both across groupings and within groupings of batterers. This diversity suggests that perfect, or even excellent, prediction will probably never be achieved. There will always be "atypical" or idiosyncratic cases that will not "fit" our prediction models. On the other hand, this finding does not negate the fact that modest prediction has already been achieved. There are distinct patterns among batterers that can be and have been identified that should inform policy and treatment.

Our conditional prediction model of multiple outcomes found that women's perceptions of risk were important predictors of repeat reassault, controlling for other significant risk markers. In addition, women's perceptions by themselves were more predictive of repeat

reassault than simulated versions of the SARA and the K-SID. Clinicians and others (e.g., de Becker, 1997; Weisz et al., 2000) have argued that women's predictions should be incorporated into risk assessment instruments. Our results support this conclusion. Moreover, proponents of SARA have recommended "enhanced clinical judgment" as a possible way to improve recidivism prediction. Perhaps women's perceptions should be a part of this enhanced judgment.

To identify the basis of women's perceptions of risk, quantitative and qualitative analysis were performed. We found that the strongest determinants of women's perceptions of risk were physical and non-physical abuse, drinking behavior, and access to the woman (based on marital status, living together, and level of contact). Women's perceptions appear to be based on rational and practical considerations, and should be incorporated into dynamic models of risk prediction and into "ongoing case management." In other words, time-varying risk markers – including women's risk appraisals – should be measured frequently to adjust assessments of the level of risk a batterer is at for repeat reassault.

LIMITATIONS OF THE CURRENT RESEARCH

As explained in the introduction, prediction of violence has been beset by several conceptual and methodological limitations. Our research attempted to address many of these by using a variety of approaches to prediction (i.e., risk markers, risk instruments, and batterer types) with conditional prediction and multiple outcomes. There remain, nonetheless, a number of other limitations. First, reliability and validity of measurement is always a concern in domestic violence research. Many of the variables are measured at the categorical level (e.g., presence or absence of previous types of abuse, heavy drinking, etc.), which may restrict prediction to some degree. Because previous violence and psychological abuse seem to be some of the strongest predictors of reassault, better measurement and calibration of past violence and psychological abuse might improve prediction somewhat. Our qualitative analysis of the

violence narratives suggests some alternative categories that might be used. Using a sequential, process model of violence, we identified an unrelenting, excessive pattern of violence in a small portion of men, who might be predicted more easily in future research. Additionally, a new coding scheme for the escalation of abuse is being developed based on the interaction pattern between partners (Eisikovits & Winstock, 2001). Women's perceptions may be predictive of violence precisely because the women can observe and monitor such escalation. Despite measurement limitations, the data set we use is one of the best databases available with regard to measurement, owing to the extent and variety of variables, the combination of quantitative and qualitative data, and the short duration between interviews. The data are much of what is practically obtainable by clinicians and what is most relevant to them. More sophisticated measurement would mean developing instrumentation and testing that is beyond the current state in the criminal justice system.

The quality and comprehensive nature of the dataset also offset, to some extent, the limitations that the risk instruments (DAS, SARA, and K-SID) were simulated using the data rather than administered in actual clinical practice. Our dataset probably has less missing data than most "real world" administrations of risk instruments. We contend that our "simulated" versions of the three risk assessment instruments are reasonable proxies for the actual instruments in that they simulate actual clinical usage. More importantly, the main objective of the simulated instruments is to provide a composite of risk markers identified as useful in assessment and prediction. They are used to help further explore the possibilities of improving prediction of intervention outcomes. Can we improve our prediction of multiple outcomes with any sort of composite instrument, rather than the assorted risk markers we considered at the outset? The answer with our data is no, at least with several of the currently conceptualized risk instruments.

Despite the comprehensive nature of the data set we use, there are some potential risk markers that are not measured. For example, some experts have implicated unstable lifestyles

as predictive of reassault (Hanson & Wallace-Capretta, 2000; Gooch, 2001), which is not available in our data set. Other potentially predictive factors that are missing include recent escalation of violence, personal motivation to change, attachment style and social support (Rankin, et al., 2000).

Our formulation of batterer types also lacks some of the indicators used in the research on batterer types (Holtzworth-Munroe & Stuart, 1994; Holtzworth-Munroe, 2001) but may offer an instructive alternative. Specifically, the Holtzworth-Munroe and Stuart (1994) typology uses non-family (general) violence and severity of family violence in conjunction with selected psychological characteristics (e.g., items related to antisocial tendencies) to classify batterers into three types. The incorporation of severity of previous family violence and use of violence on non-family members may confound to some extent the issue of the predictability of batterer types, since previous violence is such a consistent predictor of future violence by itself.

Our primary measure of batterer type, on the other hand, was based on a cluster analysis of MCMI-III (psychological) data, which resulted in four types that only approximated the Holtzworth-Munroe and Stuart typology (as discussed earlier). As a second measure of type, we used MCMI-profiles that indicate primary and secondary psychopathy, and psychopathic tendencies (Gondolf & White, 2001). Our prediction equations separately included measures of severity of previous violence and many other risk markers. Our approach enabled us to determine whether psychological "types" or psychopathy, as separate constructs, would *improve prediction beyond measures of prior behavior*.

We found in our dataset that they do not, which is an extremely important finding. In fact, none of the Axis I and Axis II disorders, as measured by the MCMI-III, were predictive above and beyond other risk markers. While psychological variables may distinguish non-batterers from batterers in general population samples or case-control clinical studies (Schumacher, Feldbau-Kohn, Slep, & Heyman, 2001), they do not appear to be very predictive

of reassault and repeat reassault in this sample of batterer program participants, once other important risk markers are controlled.

Moreover, much previous violence prediction research that finds psychopathy to be predictive of future violence among general offenders and sex offenders has used Hare's Psychopathy Checklist, which includes many behavioral items that tap into previous violent behavior (e.g., Hart, Kropp, & Hare, 1988; Rice & Harris, 1997; Serin & Amos, 1995). The Hare checklist association with future violence may be produced primarily, or to a great extent, by the incorporation of previous "antisocial behavior" (Factor 2 items on the PCL) into the instrument (Skeem & Mulvey, 2001). Similarly, the Holtzworth-Munroe and Stuart typology, to the extent that it would be predictive of future reassault, is probably also predictive because it uses violence severity as a primary indicator of batterer type.

We should also note that it may have been difficult to find an interaction between batterer type and program type because of the way we constructed batterer types. In addition, the two instructional programs do have many similarities to the two discussion-oriented programs. Therefore, the distinctions between the programs may not have been sufficient to produce different outcomes for different types of batterers.

We also acknowledge that our conditional prediction model (predicting outcomes that occur between 3 and 15 months) is not a true dynamic model in the sense of predicting short-term risk (e.g., over a 3 month time period), as can be done with Generalized Estimating Equations (GEE) models or proportional hazards models (see Jones & Gondolf, 2001). The conditional variables were based on only the intake to 3-month follow-up interval, rather than for each of the five follow-up intervals (because the incidence of repeat reassault was too low to predict in each follow-up period).

An additional limitation is the clinical nature of our sample. The sample consists of only men who were enrolled in batterer programs and not all the men who enter the courts. Our prediction, therefore, is for men enrolled in programs only. The courts are concerned about

identifying high-risk men among all of those men entering the court and not just the select group sent to batterer programs. The objective is to make referrals and sentencing appropriate to the risk. Ideally, prediction should be extended to all men who enter the courts.

Similarly, the prediction models apply to *batterer program outcome*, and **not** the *potentiality of reassault*. That is, the men's potential for reassault is mediated by the batterer programs and additional services and interventions in these four cities. Some men for whom the risk markers would have predicted repeat reassault were contained by the system. Weaker intervention systems may have produced different predictors for program outcome because more men may have been able to act on their "potential" for violence. We would argue, however, that prediction of program outcomes, rather than potentiality, is more relevant to policy development. We want to know which men are not being contained by batterer counseling, who need additional or alternative interventions.

Finally, we note that there are other appropriate analytical methods that could be applied to our data, such as latent growth curve modeling or Generalized Estimating Equations (GEE) with multiple outcomes. Latent growth curve modeling might identify batterer types based on behaviors and experiences that occur over time, whereas the current typologies are based on cross-sectional data. GEE modeling with multiple outcomes would combine the utility of dynamic modeling with the power of using multiple outcomes as opposed to dichotomous reassault outcomes. The database is very rich and other multivariate statistical techniques might be used with the data in future research. This caveat does not obviate, however, the value and utility of the current analyses. The main objective here was to test for improved prediction using multiple outcomes, and multinomial logit is the most appropriate way to do that.

IMPLICATIONS FOR RESEARCH

Our predictive study demonstrates the importance of considering multiple outcomes in batterer research. Multiple outcomes not only modestly improve prediction, but they also

expose different sets of predictors than dichotomous outcomes. Our findings suggest one reason why causal research has produced inconsistent results (see Aldarondo & Sugarman, 1996). ("Casual research" refers to studies identifying factors that help to "explain" future reassault, as opposed to simply identifying who is most likely to re-abuse.) Different predictors for repeat reassault versus one-time reassault could cancel themselves out in an equation with a dichotomous outcome. Moreover, samples with fewer men in the repeat reassault category are likely to produce different predictors than samples with more repeaters. Excellent prediction can be derived with small samples, but such prediction is generally not replicable across samples because of variations in the influential sub-categories of reabuse and reassault.

Our findings raise some questions for future research with multiple outcomes. It appears that additional variables modestly improve the prediction of multiple outcomes. However, it is uncertain how to more substantially improve prediction, or if it can be improved much further at all. One possibility is that better measurement of the existing variables and identifying additional influential variables (such as motivation) could improve prediction. The increased complexity, however, makes it more difficult to translate prediction into clinical practice. We also need further verification of risk instruments using our approach, since we were only able to simulate the risk instruments because of missing items.

We are hesitant to weigh in on the debate regarding the ultimate utility of risk prediction and risk instruments based on our findings. By using a more refined outcome (multiple outcomes) measure, we have improved prediction *in this sample*. Previous research has consistently shown that actuarial methods do predict better than clinical judgment (Grove & Meehl, 1996), so risk instruments are generally an improvement over pure clinical decision-making. Yet, much of the prediction is being driven by previous behavior, which traditionally has been used in criminal justice decision-making and victim services. Nonetheless, incorporating additional risk markers, such as men's age and women's perceptions, does improve prediction enough to warrant the continued search for additional risk markers – especially conditional and

dynamic factors. Perfect or near-perfect prediction, however, is probably beyond the reach of risk instruments. They should, therefore, be used with extreme caution and a clear understanding of their reasonable uses and limits.

IMPLICATIONS FOR PRACTICE

There remains a subjective decision about the utility of the improved prediction. The sophisticated models still only modestly predict overall outcomes and do not appear to be sufficient for clinical decisions by themselves. Even the very best prediction, achieved using risk markers among the White batterers, had a false negatives rate of 20%. In general, models that improve sensitivity, or prediction of repeat reassault, tend to have higher rates of false positives (i.e., men who are predicted to repeatedly reassault who do not). From a policy perspective, however, it could be argued that reducing false negatives (i.e., men who are predicted not to repeatedly reassault but who do) is the most important goal of risk assessment. It is probably better to intervene more heavily with false positives (i.e., give them stricter sanctions) than to not intervene enough with false negatives (i.e., give weak or no sanctions to men who are actually at high risk to reassault).

Our findings raise other implications for clinical assessment of batterers, and particularly the effort to identify and contain the most dangerous men. First, our results indicate the importance of distinguishing between "one-time reassault" and "repeat reassault" when attempting to identify "high-risk" batterers. The two outcomes have different risk markers and may not be as readily identified if combined into one group. Further, predicting repeat reassault (and potentially serious offenders) appears to be easier than predicting "any reassault." Second, the emphasis on personality traits and personality types failed to improve prediction of repeat reassault. Once other significant predictors were controlled, personality and mental health variables were either not predictive or were weak predictors of reassault outcomes. Therefore, the use of psychological assessments for identifying the extent of intervention or

level of constraint may not be that useful, especially when equal or better prediction can be achieved with measures that are much easier to obtain.

Third, the simulated risk assessment instruments appeared to offer only modest prediction in our study, and should be used with caution by batterer programs and the criminal justice system, as previous research has already recommended (see Roehl & Guertin, 2000). Results are improved somewhat with additional items and women's perceptions, reinforcing the importance of using instrument results in combination with a variety of other sources of information. Fourth, the predictive power of women's perceptions suggests the importance of obtaining and heeding women's appraisal of their situation, as advocates have long argued. Batterer program staff and the courts may have to work more closely with women's program staff to obtain such information and incorporate it into their assessments.

The quantitative and qualitative findings contradict, however, overgeneralizations about "high risk" batterers. They are not readily or easily identifiable or "typed." Many of the men who repeatedly reassault appear, according to their personality profiles, to be appropriate candidates for conventional batterer counseling (White & Gondolf, 2000; Gondolf & White, 2001). The findings also imply conditional variables enhance prediction beyond personality factors. Shifting attention from intake assessment to on-going risk management would likely improve identification and containment of the most dangerous men. Furthermore, our analysis, particularly of violent incidents, suggests that women's characteristics (i.e., levels of assertiveness, help-seeking, satisfaction with services) warrant further consideration. Prevention efforts need to consider support and safety planning with the women, as much as containment and restraint of the men.

In sum, improvement of identification and containment of the most dangerous men requires not only further differentiation of batterers (i.e., identifying the small core of excessive, unrelenting batterers), but also consideration of a wide range of information, sources, and time

frames. Conventional intake assessment or risk instruments have limited predictive power, and even the more extensive and sophisticated predictions are not particularly strong.

REFERENCES

- Akers, R. (1997). Criminological theories: Introduction and evaluation. LA. CA: Roxbury.
- Aldarondo, E., & Sugarman, D. (1996). Risk marker analysis of the cessation and persistence of wife assault. Journal of Consulting and Clinical Psychology, 64, 1010-1019.
- Aldrich, J., & Nelson, F. (1984). Linear probability, logit, and probit models. Beverly Hills, CA: Sage.
- Babcock, J., Green, C., & Robie, C. (in press). Does batterer's treatment work? A meta-analytic review of domestic violence treatment outcome research. Journal of Family Psychology.
- Berkowitz, L. (1993). Aggression: Its causes, consequences, and control. Philadelphia, PA: Temple University Press.
- Beutler, L. E., & Clarkin, J. (1990). Systematic treatment selection: Toward targeted therapeutic interventions. New York: Brunner/Mazel.
- Blackburn, R. (1996). Replicated personality disorder clusters among mentally disordered offenders and their relation to dimensions of personality. Journal of Personality Disorders, 10, 68-81.
- Blackburn, R. (1998). Psychopathy and the contribution of personality to violence. In T. Millon, E. Simonsen, M. Birket-Smith, & R. Davis (Eds.), Psychopathy: Antisocial, criminal, and violent behavior (pp. 50-68). New York: Guilford.
- Bodnarchuk, M., & Dutton, D. (1997). Wife assault treatment and criminal recidivism: An 11-year follow-up. International Journal of Offender Therapy & Comparative Criminology, 41, 9-24.
- Campbell, J. (1995). Prediction of homicide of an by battered women, In J. Cambell (Ed.) Assessing dangerousness: Violence by sexual offenders, batterers, and child abusers (pp. 96-113). Thousand Oaks, CA: Sage.

- Cano, A. & Vivian, D. (2001). Life stressors and husband-to-wife violence. Aggression and Violent Behavior, 6, 459-480.
- Chang, C., & Tuckman, H. (1989). Teacher training and teacher optimism about the American economic system. Journal of Economic Education, (Fall), 335-345.
- Cleckley, H. (1988). The mask of sanity : An attempt to clarify some issues about the so called psychopathic personality, Fifth Edition. New York: Emily S. Cleckley.
- Cunningham, M., & Reidy, T. (1998). Antisocial personality disorder and psychopathy: diagnostic dilemmas in classifying patterns of antisocial behavior in sentencing evaluations. Behavioral Sciences and the Law, 16, 333-351.
- Davis, R., & Taylor, B. (1999). Does batterer treatment reduce violence? A synthesis of the literature. Women and Criminal Justice, 10, 69-93.
- De Becker, G. (1997). The gift of fear: Survival signals that protect us from violence. New York: Dell.
- DeMaris, A. (1995). A tutorial in logistic regression. Journal of Marriage and the Family, 57, 956-968.
- DeMaris, A., & Jackson, J.K. (1987). Batterers' reports of recidivism after counseling. Social Casework, 68, 458-465.
- Dobash, R. E., & Dobash, R. P. (2000). Evaluating criminal justice interventions for domestic violence. Crime and Delinquency, 46, 252-271.
- Dobash, R. E., Dobash, R. P., Cavanagh, K., & Lewis, R. (2000). Changing Violent Men. Thousand Oaks, CA: Sage Publications.
- Dutton, D. (1986). The outcome of court-mandated treatment for wife assault: A quasi-experimental evaluation. Violence and Victims, 1, 163-175.
- Dutton, D. (1994). The origin and structure of the abusive personality. Journal of Personality Disorders, 8, 181-191.

- Dutton, D. (1998). The abusive personality: Violence and control in intimate relationships. New York: Guilford.
- Dutton, D., & Golant, S. (1995). The batterer: A psychological profile. New York: Basic Books.
- Dutton, D., & Kropp, P. (2000). A review of domestic violence risk instruments. Trauma, Violence & Abuse, 1, 171-181.
- Dutton, D., & Starzomski, A. (1993). Borderline personality in perpetrators of psychological and physical abuse. Violence and Victims, 8, 327-337.
- Dutton, D., Bodnarchuk, M., Kropp, R., Hart, S., & Ogloff, J. (1997). Wife assault treatment and criminal recidivism: An 11-year follow-up. International Journal of Offender Therapy and Comparative Criminology, 41, 9-23.
- Edleson, J., & Syers M. (1990). The relative effectiveness of group treatments for men who batter. Social Work Research and Abstracts, 26, 10-17.
- Eisikovits, Z., & Winstock, Z. (2001). Comparing men's and women's perspectives on escalation to intimate violence. Paper presented at the 7th International Family Violence Research Conference, University of New Hampshire, Portsmouth, NH, July 22-July 25.
- Fein, R., Vossekuil, B., & Holden, G. (1995) Threat assessment: An approach to prevent targeted violence. National Institute of Justice Research in Action (July), 1-7.
- Follingstad, D., Rutledge, L., Berg, B., Hause, E., & Polek, D. (1990). The role of emotional abuse in physically abusive relationships. Journal of Family Violence, 5, 107-121.
- Geffner, R. (1995). Of mice, humans, and family violence: The need for assessment-based intervention. Family Violence and Sexual Assault Bulletin, 11(3-4), 3.
- Gelles, R. (1999). Male offenders: Our understanding from the data. In M. Harway & J. O'Neil (Eds.), What causes men's violence against women? (pp. 36-48). Thousand Oaks, California: Sage.
- Gelles, R., & Tolman, R. (1998). The Kingston Screening Instrument for Domestic Violence (K-SID). University of Rhode Island, Providence, RI.

- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works. Criminology, 34, 575-596.
- Goldkamp, J. (1996). The role of drug and alcohol abuse in domestic violence and its treatment: Dade County's domestic violence court experiment. National Institute of Justice, U.S. Department of Justice, Washington, DC.
- Gondolf, E. (1988). Who are these guys? Toward a behavioral typology of batterers. Violence and Victims, 3, 187-203.
- Gondolf, E. (1997a). Expanding batterer program evaluations. In G. K. Kaufman & J. Jasinski (Eds.), Out of darkness: Contemporary research perspectives on family violence (pp. 208-218). Thousand Oaks, CA; Sage.
- Gondolf, E. (1997b). Patterns of reassault in batterer programs. Violence and Victims, 12, 373-387.
- Gondolf, E. (1999a). A comparison of reassault rates in four batterer programs: Do court referral, program length and services matter? Journal of Interpersonal Violence, 14, 41-61.
- Gondolf, E. (1999b). Characteristics of court-mandated batterers in four cities. Violence Against Women, 5, 1277-1293.
- Gondolf, E. (1999c). MCMI results for batterer program participants in four cities: Less "pathological" than expected. Journal of Family Violence, 14, 1-17.
- Gondolf, E. (2000). Reassault at 30 months after batterer program intake. International Journal of Offender Therapy and Comparative Criminology, 44, 111-128.
- Gondolf, E. (2001). Batterer intervention systems: Issues, outcomes, and recommendations. Thousand Oaks, CA: Sage Publications.
- Gondolf, E., & Beeman, A. (under review). Are men who repeatedly reassault distinguished by a special kind of violence? Violence Against Women.

- Gondolf, E., & Fisher, E. (1988). Battered women as survivors: An alternative to treating learned helplessness. Lexington, MA: Lexington Books.
- Gondolf, E., & White, R. (2000). "Consumer" recommendations for batterer programs. Violence Against Women, 6, 196-215.
- Gondolf, E., & White, R. (2001). Batterer program participants who repeatedly reassault: Psychopathic tendencies and other disorders. Journal of Interpersonal Violence, 16, 361-380.
- Gondolf, E., & Williams, O. (in press). Culturally-focused batterer counseling for African-American men. Trauma, Violence, and Abuse.
- Gondolf, E., Heckert, A., & Kimmel, C. (in press). Non-physical abuse among batterer program participants. Journal of Family Violence.
- Gooch, M. (2001). Personal communication, Portsmouth, NH.
- Goodman, L., Dutton, M., & Bennett, L. (2000). Predicting repeat abuse among arrested batterers. Journal of Interpersonal Violence, 15, 63-74.
- Gordis, E. (1997). Patient-treatment matching--A commentary. Alcohol Alert (NIAAA Bulletin), 36 (April), 3.
- Grove, W., & Meehl, P. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. Psychology, Public Policy, and Law, 2, 293-323.
- Hamberger, L.K., & Hastings, J. E. (1988). Skills training for treatment of spouse abusers: An outcome study. Journal of Family Violence, 3, 121-130.
- Hamberger, L.K., & Hastings, J.E. (1990). Recidivism following spouse abatement counseling: Treatment program implications. Violence and Victims, 5, 157-170.
- Hanson, R.K., & Wallace-Capretta, S. (2000). Predicting recidivism among male batterers. User Report 2000-06. Ottawa: Department of the Solicitor General of Canada.

- Hare, R. D. (1996). Psychopathy: A clinical construct whose time has come. Criminal Justice and Behavior, 23, 25-54.
- Hart, B. (1994). Lethality and dangerousness assessments. Violence Updates, 4(10), 7-10.
- Hart, B. (1995). Coordinated community approaches to domestic violence. Paper presented at Planning Meeting on Violence Against Women, National Institute of Justice Washington, DC, March 31.
- Hart, S., Kropp, P., & Hare, R. (1988). Performance of psychopaths following conditional release from prison. Journal of Consulting and Clinical Psychology, 56, 227-232.
- Healey, K., Smith, C., & O'Sullivan, C. (1998). Batterer intervention: Program approaches and criminal justice strategies. Washington, D.C.: The National Institute of Justice.
- Heckert, A., & Gondolf, E. (2000a). The effect of perceptions of sanctions on batterer program outcome. Journal of Research on Crime and Delinquency, 37, 369-391.
- Heckert, A., & Gondolf, E. (2000b). Assessing assault self-reports by batterer program participants and their partners. Journal of Family Violence, 15, 181-197.
- Heckert, A., & Gondolf, E. (2000c). Predictors of underreporting of male violence by batterer program participants and their partners. Journal of Family Violence, 15, 423-443.
- Heckert, A., & Gondolf, E. (2001). Factors associated with women's perceptions of safety. Paper presented at the 7th International Family Violence Research Conference, University of New Hampshire, Portsmouth, NH, July 22-July 25.
- Hilton, N.Z., Harris, G., & Rice, M. (2001). Predicting violence by serious wife assaulters. Journal of Interpersonal Violence, 16, 408-423.
- Holtzworth-Munroe, A. (2001). Examining stability and change in batterer typology. Paper presented at the 7th International Family Violence Research Conference, University of New Hampshire, Portsmouth, NH, July 22-July 25.
- Holtzworth-Munroe, A., & Stuart, G.L. (1994). Typologies of male batterers: Three subtypes and the differences among them. Psychological Bulletin, 116, 476-497.

- Hosmer, D., & Lemeshow, S. (1989). Applied Logistic Regression. NY, NY: John Wiley & Sons.
- Houghton, A. (1999). Personal communication, Denver, CO.
- Jacobson, N. J., & Gottman, J. (1998). When men batter women: New insights into ending abusive relationships. New York: Simon and Shuster.
- Jones, A., & Gondolf, E. (2001). Time-varying risk factors for re-assault by batterer program participants. Journal of Family Violence, 16, 345-359.
- Klecka, W. (1990). Discriminant Analysis. Beverly Hills, CA: Sage.
- Kropp, P., & Hart, S. (2000). The Spousal Assault Risk Assessment (SARA) guide: Reliability and validity in adult male offenders. Law and Human Behavior, 24, 101-118.
- Kropp, P., Hart, S., Webster, C., & Eaves, D. (1999). Spousal Assault Risk Assessment Guide: Users Manual. Vancouver, Canada: Multi-Health Systems Inc.
- Langford, D. (1996). Predicting unpredictability: A model of women's processes of predicting battering men's violence. Scholarly Inquiry for Nursing Practice: An International Journal, 10, 371-385.
- Lidz, C., Mulvey, E., & Gardner, W. (1993). The accuracy of predictions of violence to others. Journal of the American Medical Association, 269 1007-1011.
- Limandri, B., & Sheridan, D. (1995). Prediction of intentional interpersonal violence. In J. Campbell (Ed.), Assessing the risk of dangerousness: Potential for further violence of sexual offenders, batterers, and child abusers, (pp. 1-19). Thousand Oaks, CA: Sage.
- Margolin, G., John, R., & Foo, L. (1998). Interactive and unique risk factors for husband's emotional and physical abuse of their wives. Journal of Family Violence, 13, 315-341.
- Meloy, J. (1988). The psychopathic mind: Origins, dynamics and treatment. Northvale, N.J.: Jason Aaronson Inc.
- Millon, T. (1994). Millon Clinical Multiaxial Inventory-III manual. Minneapolis, MN: National Computer Systems.

- Millon, T., & Davis, R. (1998). Ten subtypes of psychopathy. In T. Millon, E. Simonsen, M. Birket-Smith, & R. Davis (Eds.), Psychopathy: Antisocial, criminal, and violent behavior (pp. 161-187). New York: Guilford.
- Milner, J. (2001). Risk assessment: Conceptual and measurement issues. Paper presented at the 7th International Family Violence Research Conference, University of New Hampshire, Portsmouth, NH, July 22-July 25.
- Monahan, J. (1984) The prediction of violent behavior: Toward a second generation of theory and policy. American Journal of Psychiatry, 141, 10-15.
- Monahan, J. (1996). Violence prediction: The past twenty and the next twenty years. Criminal Justice and Behavior, 23, 107-119.
- Monahan, J., & Steadman, H. (1994) (Eds.). Violence and mental disorder: Developments in risk assessment. Chicago: University of Chicago Press.
- Mulvey, E., & Lidz, C. (1985). Conditional prediction: A model for research on dangerousness to others in a new era. International Journal of Law and Psychiatry, 18, 129-143.
- Mulvey, E., & Lidz, C. (1993). Measuring patient violence in dangerousness research. Law and Human Behavior, 17, 277-278.
- National Institute of Justice (1999). Judicial oversight demonstration initiative. Washington, DC: NIJ.
- O'Leary, D. (1999). Psychological abuse: A variable deserving critical attention in domestic violence. Violence and Victims, 14, 1-21.
- Pence, E., & Paymar, M. (1993). Education groups for men who batter: The Duluth model. New York: Springer.
- Quinsey, V., Harris, G., Rice, Marnie, & Cormier, C. (1998). Violent offenders: Appraising and managing risk. Washington, DC: American Psychological Association.

- Rankin, L., Saunders, D., & Williams, R. (2000). Mediators of attachment style, social support, and sense of belonging in predicting woman abuse by African American men. Journal of Interpersonal Violence, 15, 1060-1080.
- Rice, M., & Harris, G. (1997). Cross-validation and extension of the Violence Risk Appraisal Guide for child molesters and rapists. Law and Human Behavior, 21, 231-241.
- Richards, H. J. (1993). Therapy of the substance abuse syndromes. Northvale, NJ: Jason Aronson Inc.
- Richards, H. J. (1998). Evil intent: Violence and disorders of the will. In T. Millon, E. Simonsen, M. Birket-Smith, & R. Davis (Eds.), Psychopathy: Antisocial, criminal, and violent behavior (pp. 69-94). New York: Guilford.
- Roehl, J., & Guertin, K. (2000). Intimate partner violence: The current use of risk assessments in sentencing offenders. Justice System Journal, 21, 171-197.
- Saunders, D. (1995). Prediction of wife assault. In J. Campbell (Ed.), Assessing dangerousness: Violence by sexual offenders, batterers, and child abusers (pp. 68-95). Thousand Oaks, CA: Sage.
- Saunders, D. (1996). Feminist-cognitive-behavioral and process-psychodynamic treatments for men who batter: Interaction of abuser traits and treatment models. Violence and Victims, 11, 393-414.
- Schumacher, J., Feldbau-Kohn, Slep, A., Heyman, R. (2001). Risk factors for male-to-female partner physical abuse. Aggression and Violent Behavior, 6, 281-352.
- Serin, R., & Amos, N. (1995). The role of psychopathy in the assessment of dangerousness. International Journal of Law and Psychiatry, 18, 231-238.
- Skeem, J.L., & Mulvey, E.P. (2001). Psychopathy and community violence among civil psychiatric patients: Results from the MacArthur Violence Risk Assessment Study. Journal of Consulting and Clinical Psychology, 69, 358-374.

- StataCorp. (1999). Stata Statistical Software: Release 6.0. College Station, TX: Stata Corporation.
- Steadman, H. J. (1982). A situational approach to violence. International Journal of Law and Psychiatry, 5, 171-186.
- Steadman, H., Monahan, J., Appelbaum, P., Grisso, T., Mulvey, E., Roth, L., Robbins, P., & Klassen, D. (1994). Designing a new generation of risk assessment research. In J. Monahan & H. Steadman (Eds.), Violence and mental disorder: Developments in risk assessment. (pp. 297-318) Chicago: University of Chicago Press.
- Stets, J. (1991). Psychological aggression in dating relationships: The role of interpersonal control. Journal of Family Violence, 6, 97-114.
- Straus, M. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) scales. Journal of Marriage and the Family, 41, 75-88.
- Stuart, E., & Campbell, J. (1989). Assessment of patterns of dangerousness with battered women. Issues in Mental Health Nursing, 10, 245-260.
- Sugarman, D., Aldarondo, A., & Boney-McCoy, S. (1996). Risk markers analysis of husband-to-wife violence: A continuum of aggression. Journal of Applied Social Psychology, 26, 313-337.
- Tolman, R. (1989). The development of a measure of psychological maltreatment of women by their male partners. Violence and Victims, 4, 159-177.
- Tolman, R. (1999). The validation of the Psychological Maltreatment of Women Inventory. Violence and Victims, 14, 25-38.
- Tolman, R., & Bennett, L. (1990). A review of quantitative research on men who batter. Journal of Interpersonal Violence, 5, 87-118.
- Tweed, R. G., & Dutton, D. G. (1998). A comparison of impulsive and instrumental subgroups of batterers. Violence and Victims, 13, 217-230.

Weisz, A, Tolman, R., & Saunders, D (2000.) Assessing the risk of severe domestic violence:
The importance of survivors' predictions. Journal of Interpersonal Violence, 15, 75-90.

White, R. J., & Gondolf, E. (2000). Implications of Personality Profiles for Batterers Treatment.
Journal of Interpersonal Violence, 15, 467-488.

TABLES

Table 1: Construction of Simulated Risk Instruments: Kingston Screening Instrument for Domestic Violence Offenders (K-SID)

Item	Available in the Data Set?
1. Poverty	similar
2. Age	yes
3. Drug use in past year	very similar
4. Man knows his father hit his mother	yes
5. Marital status	yes
6. Child abuse or neglect allegations	similar
7. Education	yes
8. Employment status	yes
9. Type of employment	yes
10. Previous domestic violence arrest	yes, self-reported
11. Previous violation of protection order	no

Table 2: Construction of Simulated Risk Instruments: Campbell's Danger Assessment Scale (DAS)

Item	Available in the Data Set?
1. Increased frequency of physical violence in the past year	somewhat similar
2a. Increased severity of physical violence in past year	somewhat similar
2b. Use of or threaten with weapon	yes
3. Choking or attempted choking	yes
4. Presence of gun in the house	no
5. Forced sex	yes
6. Abuser uses drugs	yes
7. Abuser threatened to kill or survivor believes abuser to be capable of killing her	yes
8. Abuser drunk almost daily	yes
9. Abuser controls partner's daily activities	very similar
10. Survivor beaten during pregnancy	no
11. Abuser violently and constantly jealous	similar
12. Survivor threatened or attempted suicide	no
13. Abuser threatened or attempted suicide	yes
14. Abuser violent toward children	yes
15. Abuser violent outside the home	similar

Table 3: Construction of Simulated Risk Instruments: Spousal Assault Risk Assessment Instrument (SARA)

Item	Available in the Data Set
<u>Part 1 (General Violence Risk Factors)</u>	
1. Past assault of family members	similar
2. Past assault of strangers or acquaintances	similar
3. Past violation of conditional release or community supervision	no
4. Recent relationship problems	at follow-up 1
5. Recent employment problems	yes
6. Victim of and/or witness to family violence as a child or adolescent	yes
7. Recent substance abuse/dependence	yes
8. Recent suicidal or homicidal ideation/intent	yes
9. Recent psychotic and/or manic symptoms	very similar
10. Personality disorder with anger, impulsivity, or behavioral instability	very similar
<u>Part 2 (Spousal Violence Risk Factors)</u>	
11. Past physical assault	yes
12. Past sexual assault or sexual jealousy	yes
13. Past use of weapons and/or credible threats of death	yes
14. Recent escalation in frequency or severity of violence	somewhat similar
15. Past violation of "no contact" orders	no
16. Extreme minimization or denial of spousal assault history	similar
17. Attitudes that support or condone spousal assault	no
18. Severe and/or sexual assault (most recent incident)	yes
19. Use of weapons and/or credible threats of death (most recent)	yes
20. Violation of "no contact" order (most recent incident)	no

**Table 4: Binary Logistic Regression for Any Re-Assault:
Intake Prediction Model (n=568)**

	Any Re-assault	
	(B)	OR
Social Background Variables		
Age		
26-35 years	-1.02***	.36
36-65 years	-1.05***	.35
Race		
African American	-.04	.96
Latino	-.89***	.41
Other	-.40	.67
Relationship characteristics		
Living with partner	.51**	1.67
Child(ren) living with batterer	-.70***	.50
Previous Behavior		
Drunk or high frequently	.55**	1.73
Verbal abuse in past 3 months	.98***	2.66
Threats in past 3 months	.70***	2.01
Non-domestic violence arrest – ever	.49**	1.64
Women's Perception of Risk		
Violence very likely	.02	1.02
Violence unlikely	.99***	2.70
Violence uncertain, don't know	.80**	2.23
Violence likely or very likely	.88***	2.43
Women's Characteristics		
Low help-seeking	-.63**	.53
Women's use of shelter	.67***	1.95
Program Characteristics		
Voluntary referral	.06	1.06
Counseling approach		
Houston (6 months, Instructional)		
Dallas (3 months, Discussion)		
Colorado (9 months, Discussion)		

% of overall cases correctly classified = 75%

% of overall re-assault cases correctly
classified (sensitivity) = 44%

*p ≤ .10 (2-tailed); **p ≤ .05; ***p ≤ .01;

**Table 5: Unordered Multinomial Logistic Regression for Multiple Outcomes: Intake Model
(n = 499)**

Risk Markers	No Abuse vs. Repeat Reassault		Control/Verbal vs. Repeat Reassault		Threats vs. Repeat Reassault		1-time vs. Repeat Reassault	
	b	OR	b	OR	b	OR	B	OR
Social Background Variables								
Age								
26-35 years	1.26***	3.54	1.02***	2.78	1.65***	5.23	.75*	2.11
36-65 years	1.19***	3.30	1.17***	3.21	1.62***	5.05	.88*	2.41
Race								
African American	.49	1.63	.86**	2.36	-.16	.85	1.01**	2.74
Latino	1.18**	3.27	1.19**	3.27	1.34**	3.82	.73	2.07
Other	1.56**	4.74	.76	2.13	.58	1.78	.31	1.36
Blue-Collar Occupation	.52	1.68	.61*	1.84	.56	1.75	.55	1.73
Relationship Characteristics								
Living with partner	-.87**	.42	-.65**	.52	-1.25***	.29	-.54	.58
Child(ren) living with batterer	.92***	2.51	1.00***	2.71	.95***	2.57	.14	1.15
Previous Behavior								
Verbal abuse in past 3 months	-1.66***	.19	-1.21**	.30	-1.47***	.23	-.65	.52
Controlling behavior in past 3 months	-1.10***	.33	-.08	.93	-.22	.80	-.75*	.47
Threats in past 3 months	-.30	.74	-.99***	.37	-.34	.71	-.06	.94
Non-domestic violence arrest - ever	-.14	.87	-.80**	.45	-.46	.63	.20	1.22
Woman Injured	-.96***	.38	-.63**	.53	-.74**	.48	-.56	.57
Women's Perceptions of Risk								
Violence very unlikely	-	-	-	-	-	-	-	-
Violence unlikely	.52	1.68	.59	1.81	.29	1.34	1.35**	3.87
Violence uncertain, don't know	-.25	.78	-1.14**	.32	-.83*	.44	.70	2.01
Violence likely or very likely	-.66	.52	-1.00*	.37	-.74	.47	-.15	.86
Woman feels very safe	-	-	-	-	-	-	-	-
Women feels somewhat safe	-1.65***	.19	-.96**	.38	-.05	.95	-1.98***	.14
Woman feels in danger, not safe, don't know	-.89	.41	-.37	.69	.01	1.01	-.23	.79
Woman's Characteristics								
Woman used shelter in past	-3.38***	.03	-1.00*	.37	-.64	.53	-2.05**	.13
Program Characteristics								
Voluntary Referral	-.25	.78	.33	1.39	.45	1.57	.60	1.83
Counseling Approach								
Houston (6 mos., instructional)	.27	1.31	-.25	.78	-.15	.86	.59	1.80
Dallas (3 mos., discussion)	-.05	.95	-.60	.55	-.71	.49	-.27	.76
Colorado (9 mos., discussion)	.25	1.29	.09	1.09	.20	1.22	.66	1.94
Model Chi-Square (df) = 270.62 (108)								
% of overall cases correctly classified = 48%								
% of Repeat Reassault cases correctly classified = 70%								
(Sensitivity)								

*p≤.10 (2-tailed); **p≤.05; ***p≤.01

Table 6: Classification Results for Unordered Multiple Logistic Regression for Multiple Outcomes: Intake Prediction Model (n=499)

Observed	Predicted					Percent Correct
	No Abuse	Control/ Verbal Abuse	Threat	One Time Reassault	Repeat Reassault	
No Abuse	50	22	12	5	8	51.5%
Control/Verbal Abuse	22	63	17	4	22	49.2%
Threats	16	28	35	3	19	34.7%
One Time Re-assault	9	17	9	11	14	18.3%
Repeat Re-Assault	9	17	7	1	79	69.9%
Overall Percentage	21.2%	29.5%	16%	4.8%	28.5%	47.7%

Sensitivity Rate for Repeat Re-Assault = 70%
False Positive Rate = 16%

Table 7: Accuracy of Predictions of Various Models for Multiple Outcomes: Intake Model (n = 499)

Variables	Sensitivity	False Positives	Overall Classification
Men's Characteristics and Reports Only	55%	20%	38%
Men's and Women's Characteristics and Reports	58%	19%	42%
Men's and Women's Characteristics and Reports, And Women's Perceptions of Risk	70%	16%	48%

**Table 8: Unordered Multinomial Logistic regression for Multiple Outcomes:
Conditional Model (m=399)**

Risk Markers	No Abuse vs Repeat Re-assault		Control/Verbal vs Repeat Re-assault		Threats vs Repeat Re-assault		One time vs Repeat Re-assault	
	B	OR	B	OR	B	OR	B	OR
Social Background Variables								
Age								
26-35 years	1.19**	3.29	.83*	2.30	1.94***	6.94	1.09*	2.9
36-65 years	.80	2.23	.67	1.95	1.35**	3.86	.78	2.1
Race								
African American	1.03*	2.81	1.18**	3.27	.70	2.01	1.32**	3.7
Latino	.74	2.10	1.72**	5.58	.94	2.56	.47	1.6
Other	.58	1.78	.35	1.42	.11	1.11	-.68	.5
Relationship Characteristics								
Living together at 3 month FU	-.76	.47	-1.53***	.22	-1.82***	.16	-.58	.56
Any relationship troubles at 3 month FU	-1.69***	.18	-.86*	.42	-.63	.53	-1.05*	.36
Antisocial Behaviors								
Drunk between intake and 3 month FU	-1.02**	.36	-.54	.58	-.21	.81	-.11	.86
Verbal abuse at 3 month FU	-1.50***	.22	-.40	.67	.25	1.29	.19	1.2
Any re-assault at 3 month FU	-1.98***	.14	-1.63***	.20	-1.61***	.20	-.46	.63
Non-DV arrest at 3 month FU	.26	1.30	-1.94	.14	-.75	.47	.78	2.1
Women's Perceptions								
Change in woman's perception of safety	.87***	2.40	.93***	2.54	.25	1.28	.41	1.5
Women's Help Seeking								
Filed protection order	-2.48**	.08	-2.18**	.11	-1.27*	.28	-2.21**	.1
Received other assistance	-.51	.60	-.52*	.60	-.28	.75	-.42	.66
Received alcohol/drug treatment	-.74	.48	-1.72**	.18	-.64	.53	-.71	.49
Man's Changes								
Man had negative change after intake	1.94**	6.97	1.55*	4.69	2.14***	8.47	2.30***	9.9
Program Characteristics								
Voluntary referral	-1.33**	.27	-.91*	.40	-.61	.55	-.21	.8
Counseling approach								
Houston (6 months, instructional)	.65	1.92	.06	1.06	1.15*	3.16	.58	1.7
Dallas (3 months, discussion)	-.45	.64	-1.11*	.33	-.39	.68	-.11	.96
Colorado (9 months, discussion)	.68	1.97	.27	1.31	.92	2.51	.31	1.3
Model Chi-square (df) = 301.72 (104)								
Cox and Snell pseudo R ² = .53								
% of overall cases correctly classified = 51%								
% of repeat re-assault cases correctly classified = 57%								

*p ≤ .10 (2-tailed); **p ≤ .05; ***p ≤ .001

Table 9: Distribution of Multiple Outcomes for Conditional Models: Women's Reports of Abuse That Occurred Between 3 and 15 Months (n = 399)

Outcome	Percentage
No physical or psychological abuse	27
Controlling behavior/verbal abuse only	27
Threats with no physical assault	20
One-time reassault	12
Repeat reassault	15

Table 10: Classification Results for Conditional Prediction Model (n=399)

Predicted						
Observed	No Abuse	Control/ Verbal Abuse	Threat	One Time Reassault	Repeat Reassault	Percent Correct
No Abuse	72	27	6	2	2	66.1%
Control/Verbal Abuse	24	52	20	2	8	49.1%
Threats	14	15	38	3	9	48.1%
One Time Re-assault	11	10	8	10	8	21.3%
Repeat Re-Assault	6	10	6	3	33	56.9%
Overall Percentage	31.8%	28.6%	19.5%	5.0%	15.0%	47.7%

Sensitivity Rate for Repeat Re-Assault = 57%

False Positive Rate = 8%

Table 11: Accuracy of Prediction of Simulated Risk Assessment Instruments and of Women's Perceptions of Risk

Variables	Dichotomous Outcome Any Reassault			Multiple Outcomes Repeat Reassault		
	Sensitivity	False		Sensitivity	False	
		Positives	Overall		Positives	Overall
K-SID, Total Score	0%	0%	67%	11%	5%	27%
K-SID, Risk Groups	0%	0%	67%	11%	5%	28%
SARA, Total Score	10%	5%	67%	43%	27%	32%
SARA, Number of Factors	7%	2%	68%	35%	23%	28%
SARA, High Risk Cut Point	0%	0%	67%	27%	16%	28%
DAS, Women's Reports	22%	8%	68%	66%	33%	33%
DAS, Combined Reports	19%	7%	69%	60%	31%	33%
Women Perceive Violence As Likely	20%	9%	67%	52%	26%	32%
Women's Perceptions of Safety	16%	9%	66%	63%	40%	31%
DAS & Violence Likely	30%	10%	70%	63%	27%	35%
DAS & Safety	24%	8%	69%	64%	28%	37%
SARA & Violence Likely	25%	8%	69%	54%	26%	34%
SARA & Safety	16%	5%	68%	55%	31%	34%
DAS, SARA, & Likely	30%	10%	70%	55%	31%	34%
DAS, SARA, & Safety	22%	8%	68%	60%	27%	37%
DAS, SARA, Likely, Safety	30%	9%	70%	52%	21%	37%
Any Psychopathic Tendencies	0%	0%	67%	54%	17%	26%

**Table 12: Unordered Multinomial Logistic regression for Multiple Outcomes:
Intake Model for African Americans (n=162)**

Risk Markers	No Abuse vs Repeat Reassault		Control/Verbal vs Repeat Reassault		Threats vs Repeat Reassault		One time vs Repeat Reassault	
	B	OR	B	OR	B	OR	B	OR
Social Background Variables								
Age								
26-35 years	1.19	3.30	.63	1.88	1.37	3.92	-.15	.86
36-65 years	1.68**	5.36	1.06	2.88	1.69*	5.41	.33	1.3
Employment Status								
Part time	-1.06	.35	.42	1.52	1.02	2.78	.33	1.4
Unemployed	.25	1.29	.33	1.39	-2.10*	.12	-.07	.93
Relationship Characteristics								
Living together	.28	1.33	.66	1.94	-.44	.64	1.55**	4.7
Prior Behavior								
Threats in past 3 months	-1.62**	.20	-2.18	.11	-1.23	.29	-1.81**	.16
Elevated MAST score (>4)	-.81	.44	-.86	.42	.68	1.97	-1.24**	.29
Psychological Characteristics								
Narcissism and antisocial	.94	2.55	1.82*	6.19	2.29**	9.90	-.08	.92
Women's Perceptions								
Violence likely, very likely, don't know	-.96	.38	-.19	.82	-.07	.94	.58	1.7
Violence unlikely	-1.82**	.16	-2.39***	.09	-1.05	.35	.49	1.6
Violence very unlikely	-1.75**	.17	-1.77**	.17	-1.41	.24	-1.06	.33
Women's Characteristics								
Use of shelter / social services	-.46	.63	-.41	.66	-.74	.48	-.59	.55
Program Characteristics								
Voluntary referral	-.43	.65	-.01	.99	.23	1.26	1.29	3.6

Model $X^2 = 92.1$ (df = 52)

% of overall cases correctly classified = 54%

% of repeat reassault cases correctly classified = 66%

*p ≤ .10 (2-tailed); **p ≤ .05; ***p ≤ .001

Table 13: Unordered Multinomial Logistic Regression for Multiple Outcomes: Intake Model For Whites (n = 216)

Risk Markers	No Abuse vs. Repeat Reassault		Control/Verbal vs. Repeat Reassault		Threats vs. Repeat Reassault		1-time vs. Repeat Reassault	
	b	OR	b	OR	b	OR	b	OR
Social Background Variables								
Age								
26-35 years	.87	2.40	.31	1.36	1.16*	3.19	-.09	.91
36-65 years	-.71	.49	.17	1.18	1.02	2.78	.33	1.39
Education								
High school	1.34*	3.80	.38	1.46	1.26**	3.54	1.98**	7.22
Some college	2.30***	9.98	.60	1.82	1.10	3.00	2.49**	12.1
College or more	2.58**	13.24	1.10	3.02	1.74**	5.70	3.58***	35.9
Relationship Characteristics								
Living with partner	-1.92***	.15	-1.28**	.28	-2.03***	.13	-2.31***	.10
Child(ren) living with batterer	1.77***	5.86	1.57***	4.80	1.55***	4.69	.70	2.02
Previous Behavior								
Controlling behavior in past 3 months	-1.96***	.14	-.63	.53	-.85*	.43	-.38	.68
Severe violence - ever	-.92	.40	-.50	.60	-1.20*	.30	-.33	.72
Alcohol related fight - ever	.78	2.17	-.85*	.43	-.45	.64	2.08	8.00
Psychological Characteristics								
Man either antisocial or narcissistic	-.54	.58	.17	1.18	.43	1.54	-1.59**	.20
Women's Perceptions of Risk								
Woman feels very safe	-	-	-	-	-	-	-	-
Women feels somewhat safe, doesn't know, feels not safe or in danger	-1.42**	.24	-.90	.41	-.07	.93	-.98	.38
Woman's Characteristics								
Has not received any previous treatment	2.55**	12.78	1.69	5.41	.36	1.44	3.11**	22.4
Woman - any contact with DV agency	.36	1.44	.47	1.60	.94	2.57	.72	2.05
Program Characteristics								
Counseling Approach								
Houston (6 mos., instructional)	1.06	1.89	-.03	.97	.34	1.40	1.99**	7.31
Dallas (3 mos., discussion)	-.49	.62	-.50	.61	-1.19*	.30	-.37	.69
Colorado (9 mos., discussion)	.63	1.89	-.11	.90	-.06	.94	.42	1.53
Model Chi-Square (df) = 148.66 (72)								
% of overall cases correctly classified = 55%								
% of Repeat Reassault cases correctly classified = 79%								
(Sensitivity)								

*p≤.10 (2-tailed); **p≤.05; ***p≤.01

Table 14: Classification Results for Intake Models: By Race

For Whites		Predicted				Percent Correct
Observed	No Abuse	Control/ Verbal Abuse	Threats	One-Time Reassault	Repeat Reassault	
No Abuse	20	4	7	1	8	50.0%
Control/ Verbal Abuse	6	20	9	2	9	43.5%
Threats	7	2	29	5	11	53.7%
One-Time Reassault	4	2	5	6	5	27.3%
Repeat Reassault	4	4	5	0	48	78.7%
Overall Percentage	18.4%	14.3%	24.7%	6.3%	36.3%	55.2%

Sensitivity = 79%; False Positive Rate = 20%

For African Americans		Predicted				Percent Correct
Observed	No Abuse	Control/ Verbal Abuse	Threats	One-Time Reassault	Repeat Reassault	
No Abuse	6	17	1	2	3	20.7%
Control/ Verbal Abuse	2	44	3	2	6	77.2%
Threats	2	10	6	1	1	30.0%
One-Time Reassault	1	7	1	10	5	27.3%
Repeat Reassault	1	8	1	1	21	65.6%
Overall Percentage	7.4%	28.6%	53.1%	7.4%	9.9%	53.7%

Sensitivity = 66%; False Positive Rate = 11.5%

For Latinos		Predicted		Percent
Observed	No	Any		
	Reassault	Reassault	Correct	
No Reassault	63	7	90.0%	
Any Reassault	10	11	52.4%	
Overall Percentage	80.2%	19.8%	81.3%	

Sensitivity = 52%; False Positive Rate = 10%

**Table 15: Binary Logistic Regression for Any Re-Assault:
Intake Prediction Model for Latinos (n=92)**

Risk Markers	Any Re-assault (B)	OR
Social Background Variables		
Age		
26-35 years	-1.33**	.27
36-65 years	-3.41***	.03
Reported parents hit one another	.68	1.97
Previous Behavior		
Verbal Abuse in past 3 months	1.43	4.16
Threats in past 3 months	1.07	2.93
Program Characteristics		
Voluntary referral	.16	1.18

% of overall cases correctly classified = 81%

% of overall re-assault cases correctly
classified (sensitivity) = 52%

*p ≤ .10 (2-tailed); **p ≤ .05; ***p ≤ .01;

Table 16: Accuracy of Prediction of Simulated Risk Assessment Instruments and of Women's Perceptions of Risk

Variables	Dichotomous Outcome Any Reassault			Multiple Outcomes Repeat Reassault		
	False			False		
	Sensitivity	Positives	Overall	Sensitivity	Positives	Overall
For Whites						
K-SID, Total Score	3%	2%	65%	63%	47%	27%
SARA, Total Score	18%	6%	67%	66%	44%	33%
DAS, Women's Reports	25%	12%	66%	75%	52%	34%
Women Perceive Violence as Likely	17%	9%	65%	47%	26%	29%
Women's Perceptions of Safety	16%	9%	65%	55%	34%	31%
For African Americans						
K-SID, Total Score	0%	0%	65%	0%	0%	35%
SARA, Total Score	10%	8%	63%	22%	13%	36%
DAS, Women's Reports	23%	9%	65%	63%	19%	42%
Women Perceive Violence as Likely	53%	20%	71%	52%	27%	40%
Women's Perceptions of Safety	18%	7%	66%	15%	10%	34%
For Latinos						
K-SID, Total Score	0%	0%	77%	0%	0%	27%
SARA, Total Score	9%	1%	78%	21%	7%	28%
DAS, Women's Reports	10%	0%	76%	46%	10%	34%
Women Perceive Violence as Likely	0%	0%	70%	69%	30%	31%
Women's Perceptions of Safety	0%	0%	70%	0%	0%	31%

Table 17: Summary of MCMI Profile Classifications by Re-assault

Classification/Grouping	Re-assault			Total %(n)
	None %(n)	Once %(n)	Repeated %(n)	
Personality Dysfunction				
Low	59 (230)	59 (40)	60 (72)	59 (342)
Moderate	26 (103)	22 (15)	22 (27)	25 (145)
Severe	15 (58)	19 (13)	18 (22)	16 (93)
Psychopathy				
Primary Psychopathic Disorder	8 (33)	9 (6)	11 (13)	9 (52)
Secondary Psychopathic Disorder	10 (39)	9 (6)	15 (18)	11 (63)
Any Psychopathic Tendencies*	39 (152)	35 (24)	54 (65)	42 (241)
Personality Continuum*				
Narcissistic-Paranoid	57 (224)	46 (31)	51 (62)	55 (317)
Avoidant-Borderline	25 (96)	35 (24)	41 (49)	29 (169)
Atypical	18 (71)	19 (13)	8 (10)	16 (94)
* $p \leq .05$				$n = 580$

Table 18: Logistic Regression of Predictors of Perceptions at 3 months of Likelihood of Jail

<i>Predictor Variables</i>	<i>For Drop Out</i>		<i>For Reassault</i>	
	(B)	O.R.	(B)	O.R.
Social Background Variables				
Underemployment	.56**	1.74	----	----
Relationship Characteristics				
At least one child living with man	.64**	1.89	.42*	1.52
Prior Behaviors				
Previously arrested (not DV)	----	----	.40**	1.49
Alcohol and/or drug treatment	.65**	1.93	.85**	2.35
Personality Traits (MCMI)				
Antisocial personality	-.73**	.48	----	----
Program Characteristics				
Voluntary program participant	-1.12**	.33	.50*	1.65
Program site				
Denver	-.007	----	-.13	----
Houston	-.66*	.52	-.61**	.54
Dallas	-.93**	.39	-.49*	.61
Dropped out by 3 months	-.41	----	-.55**	.57
Reassault by 3 month follow-up	.30	----	-.002	----
Model Chi Square (df)	75.8** (15)		39.1** (11)	
N	408		463	

NOTE: OR = Odds Ratio; MCMI = Millon Clinical Multiaxial Inventory, Version III.

*p < .05 (one-tailed); **p < .01 (one-tailed).

Table 19: Logistic Regressions of the Effect of Perceptions on Dropout and Reassault between 3 and 15 Months Controlling for Significant Predictors

	<u>Dropout</u> <u>(n=159)</u>		<u>Reassault</u> <u>(n=395)</u>	
	(B)	(B)	(B)	(B)
Perceptions of Sanctions				
Certainty				
Jail unlikely if reassault	----	----	----	----
Jail uncertain if reassault	.67 ^{ns}	----	.35 ^{ns}	----
Jail likely if reassault	.59 ^{ns}	----	.08 ^{ns}	----
Increment in X ²	1.03 ^{ns}	----	0.5 ^{ns}	----
Severity				
No sanctions if reassault	----	----	----	----
Extra legal sanctions	----	-.28 ^{ns}	----	.39 ^{ns}
Preliminary legal steps	----	-1.10 ^{ns}	----	.63 ^{ns}
Legal sanctions	----	-.46 ^{ns}	----	.16 ^{ns}
Increment in X ²	----	1.61 ^{ns}	----	1.67 ^{ns}

NOTE: *ns* = not significant.

Table 20: Logistic Regressions for the Non-Physical Abuse Variables

	EQ 1: Verbal		EQ 2: Control		EQ 3: Threats		EQ 4: Combined	
	B	OR	B	OR	B	OR	B	OR
Demographics								
Age								
26-35 years	-	-	-	-	-	-	-	-
36-65 years	-	-	-	-	-	-	-	-
White	-	-	-.57**	.57	-	-	-.40	.67**
Education								
High school	-	-	-	-	-	-	-	-
Some college or more	-	-	-	-	-	-	-	-
Full-time employment	-	-	-	-	-	-	-	-
White collar occupation	-	-	.46*	1.58	-	-	.60**	1.81
Relationship Status								
Married	-	-	.38**	1.46	-	-	.48**	1.61
Living with partner	-	-	-	-	-	-	-	-
Children living with batterer	-	-	-	-	-	-	-.38**	.68
Family-of-Origin Abuse								
Parents hit each other	-.70**	.49	-.08	.92	-.05	.95	.17	1.19
Parents hit batterer as child	.75**	2.11	.23	1.26	.18	1.19	.05	1.05
Parents had alcohol or drug problems	.47*	1.60	.08	1.09	.23	1.24	.10	1.10
Heavy Drinking								
MAST score > 5	.13	1.14	-.02	.98	.04	1.04	.06	1.06
Drunk in past year	.40	1.49	.35*	1.41	.05	1.05	.42*	1.51
Alcohol-related fights or arrests ever	-	-	-	-	-	-	-	-
Antisocial Behavior								
Antisocial personality (MCMI BR>74)	-	-	-	-	-	-	-	-
Previous arrest	-.38*	.68	-	-	.39**	1.48	-	-
Severe physical abuse ever in relationship	.61**	1.85	.63**	1.88	.37**	1.45	-	-
Previous Non-Physical Abuse								
Verbal abuse in past 3 months	1.17***	3.22						
Controlling behavior in past 3 months			.71***	2.03				
Threats used in past 3 months					1.29***	3.62		
Combined non-physical abuse in past 3 mos.							1.02***	2.77
Program Attending								
Counseling Approach								
Houston (6 mos., instructional)	-	-	-	-	-	-	-	-
Dallas (3 mos., discussion)	-	-	-	-	-	-	-	-
Colorado (9 mos., discussion)	-	-	-	-	-	-	-	-
Program "dropout" (less than 3 mos.)	.49*	1.63	.63***	1.88	.34*	1.41	.62**	1.86
Log likelihood		465.2		721.8		711.1		655.4
% of overall cases correctly classified		81%		65%		65%		69%
% of abuse cases correctly classified		97%		67%		73%		36%
(Sensitivity)								

*p<.10 (2-tailed); **p<.05; ***p<.01; - variable never entered the equation

TABLE 21: Incident Components by Reassault Category
(percentage for 536 incidents)

Incident Component	Reassault Category			Total
	Repeat	Once	None	
Relationship Status				
Living together	34	35	42	52
Not living together	14	22	34	34
Woman left/leaving	1	1	1	2
Man's Emotional State				
Screaming/yelling	7	6	7	7
Rage/flipped out	6	6	5	5
Mad upset	31	30	39	32
Hostile/mean	14	11	5	11
Deliberate/cold	4	4	3	4
Sad/depressed/crying	1	0	0	1
Disoriented/confused	0	0	0	0
Jealous/possessive	10	11	7	10
Controlling/coercive	20	21	22	21
Blaming/condemning	2	3	1	2
Substance Abuse				
Man used alcohol	13	15	18	15
Man intoxicated/drunk	15	14	10	13
Man on drugs/high	3	3	2	3
Woman used alcohol	3	9	10	7
Woman on drugs/high	1	0	1	1
Woman's Ending				
Response				
Capitulation/gives in	8	4	0	4
Aggression/counter	13	14	9	12
Gets him to leave	3	1	2	2
She leaves/escapes	19	17	16	18
Calls for help	22	24	41	28
Severity rating (>5)	24	21	16	20

TABLE 22: Violence Type and Incident Pattern by Reassault Category (percentage of 299 cases)

Incident Component	Reassault Category			Total
	Repeat	Once	None	
Violence Type				
Excessive	9	4	2	4
Severe/unstopped	16	11	18	15
Severe/stopped	9	18	12	13
Less severe	37	42	63	51
Mixed	30	26	0	14
Incident Pattern				
Consistent	39	20	NA	29
Decreasing	17	29	NA	23
Cycle	31	42	NA	37
Escalating	10	1	NA	5

TABLE 23: Woman's and Men's Issues by Reassault Category (percentage of 299 cases)

Incident Summary	Reassault Category			Total
	Repeat	Once	None	
WOMAN'S ISSUES				
Resigned/Submissiveness				
Low	56	74	94	79
Medium	10	4	6	6
High	9	1	0	2
Protective of Man				
Low	70	69	87	78
Medium	21	18	10	15
High	6	5	1	3
Trying to Stop or Control Violence				
Low	21	10	9	12
Medium	23	29	51	38
High	51	51	34	43
Seeking Help and Support				
Low	33	20	22	24
Medium	26	27	46	36
High	36	39	27	32
MAN'S ISSUES				
Possessiveness/Control				
Low	7	20	35	24
Medium	19	31	23	24
High	57	42	26	38
Manipulative/Deceptive				
Low	41	43	77	59
Medium	27	23	11	18
High	24	33	10	20
Anger/Temper				
Low	6	6	9	7
Medium	33	23	22	25
High	46	42	18	31
Decrease	3	21	41	26
Consistently Abusive				
Low	29	17	54	38
Medium	27	19	12	18
High	36	32	15	25
Alcohol/Drug Abuse				
Low	40	49	50	47
Medium	11	12	6	9
High	39	31	17	26

Note: Ratings of "decreased," "increased," and "mixed" were deleted from the table, since these categories generally accounted for less than 10% of the cases.

**Table 24: Frequency Distributions for Women's Perceptions of Risk at Program Intake
(in percentages)**

Variable	Distribution
Likelihood of violence in next three months	
very likely or likely	12.4
uncertain/don't know	19.6
unlikely	25.6
very unlikely	42.4
	(n = 542)
How safe woman feels at intake	
uncertain, not safe, much danger	11.1
somewhat safe	30.1
very safe	58.8
	(n = 539)

Table 25: Perception of the Likelihood of Violence by Perception of Safety at Intake (in percentages)

	Likelihood of Violence				Total
	Likely	Not sure	Unlikely	Very unlikely	
Safety					
Not safe	40.3	23.6	2.9	1.8	11.2
Somewhat safe	43.3	47.2	36.0	14.6	30.1
Very safe	16.4	29.2	61.2	83.6	58.7
n	(67)	(106)	(139)	(226)	(538)

$\chi^2 = 185.95$; (df = 6); $p = .000$

Somers' D (symmetric) = .49; $p = .000$

Table 26: Ordered Logistic Regression and OLS Regression for Woman's Perception of How Likely Man is to Use Violence Again – Cross-sectional Intake Model (n = 443)

Independent Variable	Logit Coefficient (B)	Standardized OLS Coefficient (B)
Social Background Variables		
Age		
18-25	---	---
26-35 years	.11	.04
36-65 years	.32	.06
Race		
White	---	---
African American	-.20	-.02
Latino	-.33	-.05
Other	-.08	.01
Employment Status		
Full-time	---	---
Part time	.77***	.13***
Unemployed	-.43	-.08*
Relationship characteristics		
Marital Status		
Never married	---	-.03
Married	.16	---
Separated/divorced	-.32	-.11**
Child(ren) living with batterer	-.08	-.02
Prior Behavior		
Verbal abuse in past 3 months	-.39	-.05
Controlling behavior in past 3 months	-.60***	-.14***
Threats in past 3 months	-.48**	-.13***
Ever use severe violence	-.71***	-.11**
Man frequently drunk or high	-.73***	-.16***
Psychological Characteristics		
Severe MCMI axis I or II disorder	-.29	-.05
Women's Characteristics		
Woman frequently drunk or high	.41*	.07
Program Characteristics		
Voluntary referral	-.60**	-.13***
Program Site		
Pittsburgh	---	---
Colorado	.11	.01
Houston	-.34	-.07
Dallas	-.35	-.07
Model $\chi^2 = 124.33$ (df=21), $p \leq .000$		
Cox and Snell pseudo R^2	.25	
OLS R^2		.25
* $p \leq .10$ (2-tailed); ** $p \leq .05$; *** $p \leq .01$;		

Table 27: Ordered Logistic Regression and OLS Regression for Woman's Perception of Safety – Cross-sectional Intake Model (n = 409)

Independent Variable	Logit Coefficient (B)	Standardized OLS Coefficient (B)
Social Background Variables		
Age		
18-25	---	---
26-35 years	-.24	-.02
36-65 years	-.64*	-.09
Race		
White	---	---
African American	-.30	-.03
Latino	-.47	-.05
Other	-.45	-.01
Education		
Less than high school	---	---
High school	.08	.01
Some college	-.55	-.08
College or more	-.25	-.03
Employment Status		
Full-time	---	---
Part time	.18	.01
Unemployed	-.08	-.03
Relationship characteristics		
Marital Status		
Never married	---	.11**
Married	.71**	---
Separated/divorced	-.28	.06
Living with partner	.73***	.12***
Child(ren) living with batterer	.02	---
Prior Behavior		
Verbal abuse in past 3 months	.14	.03
Controlling behavior in past 3 months	-.73***	-.13***
Threats in past 3 months	-.42	-.07
Ever use severe violence	-.80**	-.07*
Man frequently drunk or high	-.79***	-.14***
Psychological Characteristics		
Severe MCMI axis I or II disorder	-.42	-.06
Women's Characteristics		
Woman used violence	-.49*	-.06
Program Characteristics		
Voluntary referral	.18	.01
Program Site		
Pittsburgh	---	---
Colorado	-.33	-.05
Houston	-.92**	-.16***
Dallas	.29	.03
How likely is the man to use violence again	1.01***	.42***
Model $\chi^2 = 197.15$ (df=26), $p \leq .000$		
Cox and Snell pseudo R^2	.39	
OLS R^2		.40

* $p \leq .10$ (2-tailed); ** $p \leq .05$; *** $p \leq .01$;

APPENDIX A

Measures

In this appendix, we describe in detail the variables used in our study. Descriptions for some of these variables are redundant to the descriptions provided in the text of the report.

Outcome Measures

To replicate previous risk assessment research and for comparison to our multiple outcomes mode, we constructed a binary outcome reassault measure as follows. Assaults were assessed through a series of questions that included the following: (1) an open-ended question about "how was the relationship going," (2) descriptions of any conflicts and their circumstances, and (3) an inventory approximating the CTS (Straus, 1979). An assault was considered any incident that included one of the tactics on the Physical Aggression subscale of the CTS (i.e., push, shove, grab; slap; hit with a fist, bit, kick; hit with something, attempt to hit with something; choke or burn; threatened with a knife or gun; used a knife or gun; forced sex against will). The dichotomous outcome variable "any reassault" was coded 1 if any of the women (initial or new partners) reported any assault as occurring between intake and 15 months.

A second dichotomous "any reassault" variable was constructed based on women's reports between 3 months and 15 months. For this outcome variable, used to estimate the dynamic or conditional prediction model, reassaults that occurred between intake and the first 3-month follow-up were omitted. Thus, time-varying factors measured at the first follow-up could be incorporated into the risk models without concerns about temporal order.

The primary outcome in this study is a multiple outcomes variable, which was based on reports by the women, regarding the men's abusive behavior. Men were classified in the five following categories based on their partners' reports of abuse during the 15-month follow-up:

repeat reassaulters—more than one incident that included one of the tactics on the physical aggression subscale of the Conflict Tactics Scale;

one-time reassaulters: only one incident of physical aggression;

threatening reassault—no physical tactics but any threats (i.e., to hit, attack, or harm; to kill; to take away children or harm them, to kill or seriously harm other people; to kill or hurt himself);

controlling behavior or verbal abuse—no threats or physical tactics, but any controlling behaviors or verbal abuse (i.e., kept from talking on phone; kept from friends; stopped from going someplace; followed partner; kept from using family income; took or stole money from partner; swore or screamed; accused partner of being with another man; insulted or put down; threw, smashed, hit, or kicked something; destroyed property; or hurt a pet or pets);

no abuse—no reports of physical assault, threats, controlling behavior, or verbal abuse over 15 months.

To estimate a conditional prediction model, we also constructed a second multiple outcome variable using interviews starting at the 6-month follow-up (that collected information from 3-6 months after intake) through the 15-month follow-up. This outcome excluded the first 3-months after intake and allowed testing of a dynamic or conditional model of risk, by allowing the conditional variables measured at the 3 month follow-up to be entered as predictors in the multivariate equations.

It is important to mention some of the initial and new partners of the batterers in the study had missing follow-up interviews. Women were excluded from the analysis if their follow-up interviews 3, 4, and 5 (9-month, 12-month, and 15-month interviews) were missing. Thus, women's outcomes were coded if they had at least the third follow-up interview, even if the fourth and fifth interviews were missing. This decision was made to preserve as much of the sample as possible and is justified by the following. One, the majority of the reassaults occurred within the first six months; the amount of physical reassault and psychological abuse steadily

declined over time. Two, we conducted an analysis with only the 348 women who had complete follow-up data for all 5 follow-ups in which we calculated their partner's multiple outcome based on the first 3 follow-ups and cross-tabulated it with the partner's multiple outcome based on all 5 follow-ups. The Kappa for this cross-classification was a robust .75, justifying the use of incomplete cases. Nonetheless, the decision to use cases with missing women's reports from the fourth and fifth follow-ups will result in some downward misclassification. In other words, some batterers will be classified as less severe in their cumulative measure of abuse than they actually are. For example, a small percentage of the men who are classified as one-time reassaulters over 15 months, may have actually been repeat reassaulters, if women's reports had been obtained at 12 and 15 months. However, we are confident that the percentage of cases that are misclassified is rather small.

Predictors Measured at Intake

The predictor and control variables were taken from the background questionnaires administered to the men and women at program intake. There were five main types of variables measured at intake: social background characteristics; attributes of the relationship; the batterer's previous behaviors including alcohol use, extent of contact or experience with systems of social control, and physical and nonphysical abuse; batterer's dispositional or personality attributes, including measures of mental health; and women's perceptions of risk and previous help-seeking behavior and service contact. The variables we used have been identified as risk markers in previous research and/or have been discussed in the domestic violence literature as possibly being associated with assault.

The social background variables for the multivariate equations are age (1 = 18-25 years, 2 = 26-35 years, 3 = 36-65 years); race (1 = Euro-American, 2 = African American, 3 = Latino); employment status (1 = full-time, 2 = part time, 3 = unemployed); and three indicators of socioeconomic status, occupational status (0 = white collar and semi-skilled, 1 = blue collar),

education (1 = less than high school, 2 = high school, 3 = some college, 4 = college graduate or more), and the women has received welfare or public assistance in the last 3 months (0 = no, 1 = yes).

Relationship characteristics are marital status of the man at intake (1 = never married, 2 = married, 3 = separated or divorced), living situation (0 = living apart, 1 = living together), time involved with partner at intake (0 = up to three years, 1 = three or more years), and number of children living with the man (0 = none, 1 = one or more).

Several measures of the man's alcohol and drug use and contact with systems of social control were used to assess which would be the strongest predictor of reassault. Measures of alcohol and/or drug use were the following: a binary indicator of an elevated Michigan Alcohol Screening Test (MAST) score (> 4) was constructed (0 = not elevated, 1 = elevated), alcohol-related crime reported on the Michigan Alcohol Screening Test (0 = no, 1 = one or more alcohol-related fights, drunk-driving arrests, or other alcohol-related arrests in the past year), man's report of how often he was drunk or high in the last year (0 = never or almost never, 1 = at least on to two times in a three-month period to almost every day) adjusted for the woman's report of how often the man was drunk or high in the last year, man's report of alcohol and/or drug treatment (0 = none or self-help, 1 = detox and/or inpatient treatment and/or outpatient treatment). Different alcohol measures were tried because of the tendency for discrepancies between alcohol tests and self-reports. Whether the man was arrested for anything other than domestic violence in the previous year (0 = no, 1 = yes) was also computed.

The women were also asked about other forms of abuse at intake, and these reports were adjusted with men's reports of the same behaviors. They were administered an inventory of *controlling behaviors* (i.e., kept from talking on the phone; kept from spending time with friends; stopped from going some place; followed against will or knowledge; kept from using income or savings; taking money from the person), *verbal abuse* (i.e., swore or screamed at; accused of other relationship; put-down, insulted, called names), and *threats* (i.e., threatened to

hit, attack, or harm the woman; kill any person; take away or harm children; harm other people; kill or hurt self). Women's reports of the men ever committing *severe violence* prior to program intake was also included as a control variable in the prediction models, with a value of 1 indicating a "yes" response to any of the so-called severe items on the CTS (hit with a fist, bit, kick; hit or tried to hit with something; choke or burn; threatened with or used a weapon; forced sex). In addition, the women reported whether the batterer had ever caused any other injuries other than bruises (0 = no, 1 = yes), whether they had ever sought medical help (0 = no, 1 = yes), and whether they had ever been hospitalized (0 = no, 1 = yes). These last three variables are used as indicators of severe abuse. Presumably, women who were most severely abused in the incident that brought the batterer to treatment may be at greatest risk for repeated reassault.

Previous research has implicated personality disorders and psychological characteristics as determinants of intimate partner violence. Therefore, a series of variables pertaining to the man's dispositional or personality attributes were constructed using results from the Million Clinical Multiaxial Inventory, Version III (MCMI; Million, 1994). The MCMI is a 175-item test with 24 subscales that correspond to Axis II (personality disorders: e.g., antisocial or narcissistic personality) and Axis I (major disorders: e.g., major depression, schizophrenia) diagnoses. A base-rate (BR) score of 75 or above (i.e., an "elevated score") suggests that symptoms for a particular disorder are above the norm in a clinical population, or "present". Elevated scores ($BR \geq 75$) on the paranoid, borderline, schizotypal, major depression, delusional, or thought disorder subscales are considered evidence of "severe" pathology. Each personality type and/or clinical disorder was coded 0=absence and 1=presence. We also explored various combinations of MCMI subscales. For example, one combined index measured elevated scores on any severe clinical syndrome (Axis I) and another measured elevated scores on any clinical personality disorder (Axis II).

Batterer types were determined through a cluster analysis of the MCMI-III data. Four groups emerged from the cluster analysis: 1) little or no pathology, 2) narcissism or antisocial, 3) dependent/avoidant, and 4) multiple disorders. These groups approximate the prevailing batterer typologies (see Holtzworth-Munroe & Stuart, 1994), which have been presumed to be associated with reassault but have not been adequately tested as risk markers.

Women's perceptions of risk and previous help-seeking behavior and service contact were obtained through the interviews conducted with women at the time of program intake. With regard to perceptions of risk, the women were asked, "How safe do you feel at this point?" (coded 1 = uncertain, not safe, in much danger, 2 = somewhat safe, and 3 = very safe), and "How likely is it that your husband will become violent towards you during the next 3-months?" (coded 1 = very likely or likely, 2 = uncertain/don't know, 3 = unlikely, 4 = very unlikely).

Women's help seeking was measured a number of ways. First the women were asked whether they had ever gone to a shelter (0 = no, 1 = yes). Second, they were asked whether the woman had previously called the police (0 = no, 1 = yes). A third variable was summed the number of help seeking strategies the woman had pursued among the following, gone to a shelter, threatened to divorce or separate, stayed overnight at a friend's or relative's home, and contacted a counseling program or social services. A fourth variable was coded 1 if the woman had ever sought legal help by previously calling the police, obtaining a protection order, seeking legal assistance, or pressing charges against the man, and a 0 if not. A fifth variable was coded 1 if the women had ever pursued any of the above types of help, and 0 if they had not.

Previous service contact was assessed in a variety of ways. First, the women were asked if they had been contacted by staff from a women's center or shelter, legal office, or batter program staff in the past few months (0 = no, 1 = yes). Second, the women were asked if they had received any kind of treatment for alcohol or drug abuse (0 = no, 1 = yes). If they refused to answer this treatment question, a refusal variable was coded 1 (versus 0 if they answered the question). Third, the women were asked if they had received any other kinds of

help or support for any other problems other than domestic violence, such as counseling or from a mental health clinic or psychiatric hospital (0 = no, 1 = yes).

Program Variables: The effects of program context are captured by including the four sites—Pittsburgh, Denver, Houston, and Dallas—as a categorical variable. In all multivariate equations, the Pittsburgh site is the omitted reference group, because it has the greatest control through 30-day mandatory court review, the shortest duration of treatment, and the least amount of services. Thus, we will be able to ascertain how length and which additional services affect reassault outcomes, above and beyond a three-month program with few services but with mandatory court review. Court-referred (coded 0) versus voluntary status (coded 1) was determined by the referral source indicated on the background questionnaire. Court-referral included being required to attend the batterers program as a result of a criminal or civil court order or probation officer recommendation. Those who were not court-referred were attending “voluntarily” through a referral from another social service or urging from their partner. (In Pittsburgh, 5% of the men were “voluntary” and 20% were at each of the other sites.)

Predictors Measured at First 3-Month Follow-up

Conditional variables were identified from the 3-month follow-up interval with the woman. They included living arrangements, contact between partners, employment status, substance use, further arrests, and use of additional services and treatment. Time-varying predictors that were measured at both intake and follow-up 1 were employment status, living situation, child living with the batterer, use of verbal abuse, controlling behavior, threats, or physical abuse during the first three months of follow-up, arrest for non-domestic offense, the batterer had been frequently drunk or high, women’s perceptions of how likely the man was to hit again and of safety, injuries were caused by the batterer, and medical help or hospitalization was necessary owing to injuries caused by the batterer. Conditional variables that were not measured at intake are described below.

Program dropout was derived from the number of sessions the participants attended at each program according to program records. This information was converted into a dichotomous variable, coded 1 if the batterer attended at least 12 sessions and 0 if he did not. We used the 12-week threshold to accommodate the minimum required sessions of 12 weeks at two of the research sites (Pittsburgh and Dallas) and what appears to be a threshold regardless of program length (i.e., more than 90 percent of the dropouts in the longer programs occur within the first three months). Reasons for dropping out do vary. Some men were formally dismissed for reassaulting, whereas others simply withdrew voluntarily from the program.

Both the men and women were asked an open-ended question about how their relationship had been going for the last three months, followed by specific yes-no questions regarding were there any major problems or hassles, any communications problems, any conflicts or arguments, or poor sexual relations. Binary indicators were constructed separately for the men and women to measure if they perceived they had any of the preceding relationship problems (0 = none, 1 = any of the four problems listed above). Only the men were asked if there was anything that their partner had done that angered them or might have "set them off" during the previous three months, which was coded 0 for a negative response and 1 for an affirmative response. The men were also asked if they had any close calls, in which they felt like hitting their partners, which was coded 1 for an affirmative response. The "anger" and "close calls" variables are indicators of conflict in the relationship that may not be tapped by the other predictors. They may also tap into behavioral intentions or a predisposition to use violence, which the men may not be willing to directly admit to.

Binary indicators were constructed for other relationship and behavioral variables that occurred in the previous three months based on the women's reports: was the woman emotionally hurt by the man in any way, had the woman gone to the hospital, was there any time where the woman wanted to or needed to go to the doctor or hospital but had not, had the

man physically punished or hit any of the children living with the woman, and had the man been abusive to the children in any way, such as verbally or sexually abusive over the last three months.

A large number of help-seeking variables were measured at follow-up one to reflect what happened over previous 3 months. Binary indicators and/or summated scales were created separately for the men's and women's reports for the following variables: woman had any contact with a shelter; woman stayed overnight in a shelter; total amount of shelter contact; woman took some kind of legal action; woman called the police; woman filed for a protection order; woman filed for divorce; woman received other kinds of counseling or mental health treatment; woman received individual counseling; woman received family or couples counseling; total amount of counseling or mental health services; woman received any type of treatment for alcohol or drug abuse; woman used a self-help program or 12-step program for alcohol or drug treatment; total number of types of treatment received for alcohol and drug abuse; woman received any other kinds of assistance, training, or social services; woman received welfare, public assistance, or food stamps; total number of other kinds of assistance received; woman engaged in any other, informal types of responses; woman talked to a friend or relative about their problems; woman talked to clergy or church person; woman threatened divorce or separation; woman stayed with a friend or relative overnight; and total number of informal responses used.

A total quality of life indicator was computed by summing positive responses by the women to a number of quality of life measures available in the first follow-up.

Binary indicators were constructed to reflect the men's and women's perceptions of what the batterer had done to avoid incidents of violence during the past three months. Specifically, both the men and the women were asked if the man took a formal "time out" (0 = no, 1 = yes); stopped drinking or using drugs (0 = no, 1 = yes); stayed away from the woman (0 = no; 1 = yes); used any of a variety of strategies such as left the room or house, stopped

arguing/withdrew, thought before acting, took a formal time out, or used self-talk (0 = no, 1 = yes); and used any of the following strategies, discussed the problem and solutions, talked about his feelings, or called or talked to a friend, counselor, or hotline (0 = no, 1 = yes). These measures were constructed separately for the men and the women.

Binary indicators and summated scales were computed for the following responses given by the men and the women to an open-ended question about how the man had changed, if at all, in the previous three months: man was not changed; man used avoidance; man used restraint; man demonstrated behavioral change; man demonstrated relational change; man showed greater awareness; and the man demonstrated conversion. The woman was also asked if their partner now admitted that he had a problem (0 = no, 1 = yes), and the men were asked if they had changed at all during the past three months (0 = no, 1 = yes).

Binary indicators and summated scales were computed for the following responses given by the men and women to an open-ended question regarding what should be changed or improved about the batterer treatment program: don't know/uncertain, contact with the woman, counseling for the woman, change counseling methods, change the structure of the program, and create more consequences for the men for failure to attend.

Both the men and the women were asked if the man received any other assistance over the previous three months. Separate binary indicators were constructed that were coded 1 if the man received or sought any other help, assistance, or support of any kind other than the batterer treatment program. Binary indicators were also computed if the man received alcohol or drug abuse treatment; and if the man used a self-help or 12-Step program. The men and women were also asked if the man received individual counseling (0 = no, 1 = yes) or any other type of counseling other than the batterer program (0 = no, 1 = yes). Separate binary indicators were coded 1 if the man received any other assistance, such as church attendance, recreational activity, parenting program, special clubs or support groups, job training or placement program, reading self-help books or materials, medical help, or any other type of support.

Perceptions of sanctions were assessed with a measure of severity and a measure of certainty for batterer dropout and reassault. At the first 3-month follow-up, the batterers were asked two open-ended questions with regard to what would happen if they "are physically abusive again" and "if you stop attending the program or have stopped." Interviewers checked all applicable responses, which we grouped into four types of perceived sanctions: none (no sanctions or uncertain), extralegal sanctions (partner would leave, other, counselor/group would talk to man), preliminary legal steps (partner would call police, man would be arrested, sent back to court, dismissed from program), and legal sanctions (fined, sent back to program, jail). We present all that are mentioned as well as the most severe mentioned. This variable serves as our measure of severity of sanctions.

In addition, the batterers were asked how likely it was "that you'd be put in jail" for program dropout and partner reassault. The five-level Likert-type response was collapsed into (1) likely, (2) don't know/uncertain and other, and (3) unlikely categories. This variable serves as our measure of certainty of sanctions.

The perceptions of sanctions were assessed at the 3-month follow-up to capture the possible effect that program contact may have added to the perceptions beyond the batterer's arrest and court experience. In the initial program sessions at each program, staff instructed the participants about the fact that domestic violence is a crime and about the designated legal sanctions for program dropout and reassault.

One additional factor measured after intake at each follow-up interview was amount of contact. Two variables were constructed from this question. First, men who had no contact with their primary partner between the 3-month and 15-month follow-ups were coded 1, whereas men who had contact at any time during the 3- to 15-month follow-up period were coded 0. Second, men who had no contact during any of the 3-month intervals were coded 1, whereas men who had contact during each of the 3-month intervals were coded 0. Although no-contact cases have a slightly lower risk of reassault, reassaults still occur because the

batterer may pursue a woman who leaves. As Gondolf (1997b) has suggested, no contact appears difficult for women to achieve, because of issues such as child visitation, financial dependence, and pursuit by some men.

Simulated Risk Assessment Instruments

In this section, we describe briefly the three risk instruments that we were able to simulate using our data: the Kingston Screening Instrument for Domestic Violence Offenders (K-SID), the SARA (Spousal Assault Risk Assessment), and Campbell's Danger Assessment Scale (DAS). The DAS was designed to assess the potential for lethality, whereas the K-SID and the SARA were designed to assess the risk for further violence and not necessarily lethality. Tables 1, 2, and 3 show the items on each instrument and their availability in our data set. These are three of the most popular risk instruments currently in use, either formally or informally.

The K-SID was developed by Richard Gelles and is used statewide in Connecticut for sentencing in domestic violence cases (Roehl & Guertin, 2000). It has three parts: the poverty chart, a severity and injury index, and an index of ten risk markers. Total K-SID scores are classified into risk categories, although we assess how well both the total scores and the risk classifications predict reassault. A study of the K-SID is currently underway by Eleanor Lyons in Connecticut, but prediction results are not available at this time. As can be seen in Table 1, our simulated version of the K-SID is excellent. The only item that is completely missing out of the eleven items is previous violation of a protection order. We have seven of the eleven items available, and similar items for the remaining three items.

The DAS was developed by Jacqueline Campbell (1995) based on known risk markers for lethal violence. The scale consists of fifteen yes or no questions; scoring is based on the total number of yes answers. There are no cutoff scores, risk categories, or weighting of the items. Five studies have shown very good reliability and moderate construct validity (Roehl & Guertin, 2000). A study by Goodman, Dutton, and Bennett (2000) showed that the DAS was

modestly predictive of misdemeanor reassault in the short term. However, their original sample was quite small ($n = 92$) and their follow-up sample was even smaller ($n = 47$), which limits the utility of their findings. A recent study by Weisz, Tolman, and Saunders (2000) also simulated the DAS and found it to be a modest predictor of reassault at a 4-month follow-up when used in conjunction with other significant predictors. At this time, the DAS is not being formally used in sentencing (Roehl & Guertin, 2000). As Table 2 shows, our simulated version of Campbell's Danger Assessment Scale is very good. Three of the fifteen items in the index are completely missing, five of the items are somewhat to very similar, and the remaining seven items are available in our data set.

The SARA was developed at the British Columbia Institute on Family Violence (Kropp, Hart, Webster, & Eaves, 1999). It consists of a clinical checklist of twenty risk factors, which are grouped into five areas: criminal history, psychosocial adjustment, spousal abuse history, current offense characteristics, and other (Roehl & Guertin, 2000). Each of the twenty items is scored 0 to 2; risk is considered to increase with the number of items that are present (score of 2). We use total scores in this study, number of factors present, and risk categories, to see which scoring method is most predictive of reassault. The SARA is currently used by courts in two states (Colorado and Vermont) to determine level of supervision and intervention, and in Canada to determine treatment plan based on level of risk (Roehl & Guertin, 2000). It relies on victim and offender interviews, as well as criminal records. Although interrater reliability is high and internal consistency is moderate, evidence of predictive validity is modest. Only enhanced clinical judgments predict recidivism, and the prediction is modest (Kropp et al., 1999; Kropp & Hart, 2000). A large-scale study is being conducted in Colorado to assess the predictive validity of the SARA. Preliminary results suggest that the SARA predicts reoffense better than chance (Dutton & Kropp, 2000). As Table 3 reveals, the simulated version of the Spousal Assault Risk Assessment instrument is very good as well. Four of the twenty items in the index are

completely missing, six of the items are similar, and the remaining ten items are available in our data set.

Coding of Psychopathology Using the MCMI-III

For our exploration of alternate subtypes of batterers, we developed measures of primary psychopathy, secondary psychopathy, and psychopathic tendencies using MCMI-III profiles (Gondolf & White, 2001). We followed the interpretations recommended in the instrument manuals with one revision. We identified psychopathic tendencies according to profile configurations recommended by experts on psychopathy and the MCMI (Blackburn, 1998; Millon & Davis, 1998). Any evidence of psychopathic tendencies was given priority over other possible interpretations in order to ensure the maximum inclusion of such tendencies. The broader and more liberal conception of psychopathy is likely to identify more men than narrower conceptions previously used in the field. We will describe our procedures in more detail in the results section.

Qualitative Coding of Violence Narratives

We conducted qualitative case reviews of women's narratives and atypical cases to determine whether there are behaviors or situations, rather than personality types, which distinguish the repeat reassaulters. To explore for alternative abuse outcomes, our research team coded the women's descriptions of the violent incidents using a sequential, situational conception of violence (Monahan, 1996; Mulvey & Lidz, 1993). These descriptions of violence were prompted by an open-ended question that preceded the administration of the CTS, which asked the women to tell the interviewer about what happened in the most severe incident of physical abuse. The interviewers then coded the issues, circumstances, precipitants, alcohol use, man's emotional state, couple interaction, pattern of tactics, and the woman's and the man's response to the violence. The codes for the various components were, then, cross-

tabulated with the categories for reassault (no, once, repeat) to identify differences across the outcomes. The assistants also wrote their overall impressions and observations of the violence in each case, and other researchers summarized this information and used it to confirm and elaborate the cross-tabulations.

In an additional attempt to identify distinguishing characteristics of repeat reassaulters, we identified "atypical" men and women in our evaluation. Our rich database and extensive interviewing enabled us to do this. At weekly meetings of the research assistants conducting the interviews, the interviewers nominated any atypical cases among their interviewees. The "atypical cases" were considered any man, woman, or couple who were distinct, unique, or exceptional in terms of their characteristics or circumstances. The research assistants and supervising researchers then discussed the nominations, compared them to other cases and previous nominations, and agreed to classify the cases as "atypical" or not. At a later time, two research assistants, who were not involved in the interviewing, read through the fifteen months worth of interview records for each atypical case and wrote a case summary. The case summaries were then grouped according to the distinguishing features of the cases and will be discussed in the results section.

PUBLICATIONS AND PRESENTATIONS SUPPORTED BY NIJ GRANT

PUBLICATIONS

Articles in Refereed Journals

Heckert, A., & Gondolf, E. (2000). The effect of perceptions of sanctions on batterer program outcome. Journal of Research on Crime and Delinquency, 37, 369-391.

Gondolf, E., & White, R. (2001). Batterer program participants who repeatedly reassault: Psychopathic tendencies and other disorders. Journal of Interpersonal Violence, 16, 361-380.

Gondolf, E., Heckert, A., & Kimmel, C. (Forthcoming, September, 2002). Non-physical abuse among batterer program participants. Journal of Family Violence.

Gondolf, E., & Beeman, A. (under review). Are men who repeatedly reassault distinguished by a special kind of violence? Violence Against Women.

Book Chapters

Gondolf, E. (2001). Chapter 8. Batterer intervention systems: Issues, outcomes, and recommendations. Thousand Oaks, CA: Sage Publications.

Heckert, A., & Gondolf, E. (forthcoming). Predicting levels of abuse and reassault among batterer program participants. Fisher, Bonnie (Ed.), 2000 Violence Against Women and Family Violence Conference Proceedings. Department of Justice.

Papers in Progress (with working titles)

Heckert, A., & Gondolf, E. Predicting levels of abuse and reassault among batterer program participants.

Heckert, A., & Gondolf, E. The utility of simulated risk instruments in predicting repeat reassault among batterer program participants.

Heckert, A., & Gondolf, E. Predicting reassault among batterer program participants: Does race matter?

Heckert, A., & Gondolf, E. Using women's perceptions to predict reassault among batterer program participants.

Heckert, A., & Gondolf, E. Determinants of women's perceptions of risk.

PRESENTATIONS

Heckert, A., & Gondolf, E. (1999, June 11). *Predicting levels of abuse and reassault among arrested batterers*. Paper presented at the National Institute of Justice, Violence Against Women and Family Violence Research and Evaluation Program Annual Grantees Meeting. Washington, DC.

Heckert, A., & Gondolf, E. (1999, July). *The effect of perceptions of sanctions on batterer program outcomes*. Paper presented at the 6th International Family Violence Research Conference at the University of New Hampshire, Durham, NH.

Gondolf, E., & White, R. (1999, July). Batterer program participants who repeatedly reassault: Psychopathic tendencies and other disorders. Presented at the 6th International Family Violence Research Conference at the University of New Hampshire, Durham, NH.

Heckert, A., & Gondolf, E. (1999, November 19). *Predicting levels of abuse and reassault among arrested batterers*. Paper presented at the annual meeting of the American Society of Criminology, 51st Annual Meeting, Toronto, Canada.

Heckert, A., & Gondolf, E. (2000, October 2). *Predicting levels of abuse and reassault among arrested batterers: Comparing risk assessment instruments with women's perceptions*. Invited paper presented at the National Institute of Justice Research Conference on Violence Against Women and Family Violence. Washington, DC.

Heckert, A., & Gondolf, E. (2001, July). *Determinants of women's perceptions of safety and likelihood their partner will reassault*. Paper presented at the 7th International Family Violence Research Conference at Portsmouth, NH.

PROPERTY OF
National Criminal Justice Reference Service (NCJRS)
Box 6000
Rockville, MD 20849-6000

Heckert, A., & Gondolf, E. (2001, November). *Using Women's Perceptions to Predict Reassault Among Batterer Program Participants*. Paper to be presented at the annual meeting of the American Society of Criminology, 53rd Annual Meeting, Atlanta.